

Model of Brain Activation Predicts the Neural Collective Influence Map of the Brain

Flaviano Morone,¹ Kevin Roth,^{1,2} Byungjoon Min,¹ H. Eugene Stanley,³ and Hernán A. Makse^{1,*}

¹*Levich Institute and Physics Department, City College of New York, New York, NY 10031, USA*

²*Theoretical Physics, ETH Zürich, 8093 Zürich, Switzerland*

³*Center for Polymer Studies and Physics Department, Boston University, Boston, MA 02215*

Efficient complex systems have a modular structure, but modularity does not guarantee robustness, because efficiency also requires an ingenious interplay of the interacting modular components. The human brain is the elemental paradigm of an efficient robust modular system interconnected as a network of networks (NoN). Understanding the emergence of robustness in such modular architectures from the interconnections of its parts is a long-standing challenge that has concerned many scientists. Current models of dependencies in NoN inspired by the power grid express interactions among modules with fragile couplings that amplify even small shocks, thus preventing functionality. Therefore, we introduce a model of NoN to shape the pattern of brain activations to form a modular environment that is robust. The model predicts the map of neural collective influencers (NCIs) in the brain, through the optimization of the influence of the minimal set of essential nodes responsible for broadcasting information to the whole-brain NoN. Our results suggest new intervention protocols to control brain activity by targeting influential neural nodes predicted by network theory.

Experience reveals that the brain is composed of massively connected neural elements arranged in modules [1, 2] spatially distributed yet highly integrated to form a system of network of networks (NoN) [3–9]. These modules integrate in larger aggregates to ensure a high level of global communication efficiency within the overall brain network, while preserving an extraordinary robustness against malfunctioning [3–5].

The question of how these different modules integrate to preserve robustness and functionality is a central problem in systems science [3–5]. The simplest modular model [2] would assign the same function to the connections inside the modules and across the modules. However, the existence of modularity gives rise to two types of connections of intrinsically different nature: the intermodular links and intramodular links [6, 9–11]. Intramodular links define modules usually composed of clustered nodes that perform the same specific function, like for instance, the visual cortex responsible for processing visual information. Besides having intralinks, nodes in a given module may have intermodular connections to control or modulate the activity of nodes in other spatially remote modules [3, 5, 6, 9, 12].

For example, in integrative sensory processing, the intermodular links mediate the bottom-up (or stimulus-driven) processes from lower-order areas (eg, visual) to higher-order cortical ones, and top-down (or goal-directed) control from higher levels to lower ones [3, 5, 6, 12]. Indeed, in studies of attention, the pattern of brain activation indicates that high-level regions in dorsal parietal and frontal cortex are involved in controlling low-level visuo-spatial areas forming a system of networks connected through intermodular control links (dorsal-frontoparietal NoN) [6, 12]. The purpose of this work is to introduce a minimal model for a robust brain NoN

made of such intramodule connections and intermodular controllers, which, by abstracting away complexity, will allow us to make falsifiable predictions about the location of the most influential nodes in the brain NoN. Targeting these neural collective influencers (NCIs) influencers may help in designing intervention protocols to control brain activity prescribed by network theory [13, 14].

RESULTS

We consider a substrate NoN composed by two modules (Fig. 1a, below we generalize to more modules). Every node i has k_i^{in} intramodular links to nodes in the same module and k_i^{out} intermodular links to control other modules (for the sake of simplicity we first consider the case $k_i^{\text{out}} \in \{0, 1\}$ for every node i ; the general case $k_i^{\text{out}} \in \mathbb{N}_0$ will be treated later). Because controlling links connect two different modules, they are fundamentally different from intramodular ones: the latter encode only the information about *if* two nodes are connected or not inside a module, whereas the former carry the additional information about *how* nodes control each other in two different modules. We arrive to an important difference between both types of links which has been recognized in previous NoN models [10]. An intermodular link between two nodes exists because of their mutual dependence across two distinct modules performing different functions. Therefore, it is reasonable that for this intermodular link to be active, both nodes across the modules should be active. On the contrary, nodes inside a module connected only via intramodular links that do not participate in intermodular dependencies will be active independently on the other module's activity. The intralinks and interlinks are analogous to the strong and weak links defining hierarchical modules in the NoN in Refs. [9, 11].

To elaborate on the mode of intermodular control, think of a node i as a receiver of inputs external to the

* hmakse@lev.ccny.cuny.edu

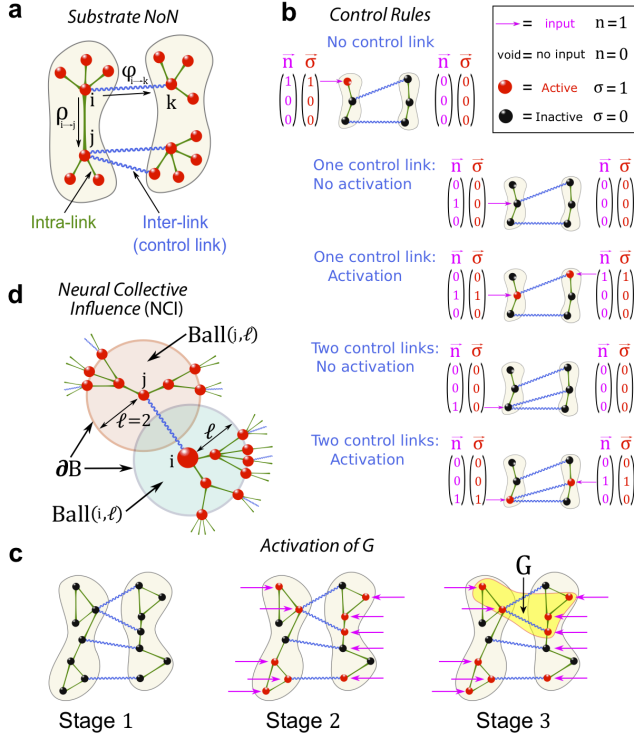


FIG. 1. Definition of NoN model. **a, Substrate NoN.** Each node has k_i^{in} intramodular links and k_i^{out} intermodular control links. Nodes send information through two messages to their neighbors: a message $\rho_{i \rightarrow j}$ along the intralink and a message $\varphi_{i \rightarrow k}$ along the control link. **b, Control rule Eq. (3).** A node i in the substrate NoN may receive an external input $n_i = 1$, or not $n_i = 0$. If the node has no control link it activates as soon as it receives the external input: $n_i = 1 = \sigma_i$. If it has 1 control link, it activates $\sigma_i = 1$ if and only if it receives the input, $n_i = 1$, and its neighbor at the edge of the interlink receives the input as well ($n_j = 1$). If it has 2 control links (or more) it activates ($\sigma_i = 1$) iff it receives the input and at least one node among its j neighbors at the edge of the interlink also receives an input $n_j = 1$, otherwise it does not activate ($\sigma_i = 0$). **c, Activation of the mutual giant component.** Global communication in the NoN is measured through the largest active component G which is measured only with the active nodes $\sigma_i = 1$. We start with a NoN with no external input (all $n_i = 0$), then $G = 0$ (Stage 1). Once an input is presented to the brain NoN (Stage 2) nodes activate according to the rules in **b**, and the largest component of activated nodes defines G (Stage 3), which it is not necessarily equal to the sum of the individual giant components of the single networks. Note the crucial ingredient of the model (not shared by the model of [10]): active nodes ($\sigma = 1$) may exist outside G , and they can have intermodular control links with other nodes outside the giant component. Thus, nodes can be active without being part of the giant component of their own network in contrast to the rules in [10]. **d, Collective Influence.** The collective influence of node i is determined by the sum of the degree of nodes in G on the surface of two balls of influence with radius ℓ : $\partial\text{Ball}(i, \ell)$ centered at i , and $\partial\text{Ball}(j, \ell)$ centered at j , where j is a neighbor of i at the edge of an interlink having out-degree $k_j^{\text{out}} = 1$.

NoN such as external sensory inputs to the primary visual cortex (Fig. 1b and SI Text). The input variable $n_i = 1, 0$ specifies whether i receives the external input ($n_i = 1$) or not ($n_i = 0$). For example, in the visual system, $n_i = 1$ is the subset of nodes receiving inputs in the earlier stages in cortical sensory processing [6].

According to the discussion above, the input n_i alone does not determine the activation/inactivation state of the node i , which we measure by the state variable σ_i taking values $\sigma_i = 1$ if i is activated, and $\sigma_i = 0$ if not. When i has a control link with j in another network, the state σ_i is determined not only by the input n_i , but also by the input n_j to j : node i is activated $\sigma_i = 1$ only when both nodes receive the input ($n_i = 1$ and $n_j = 1$). On the contrary, when at least one of the i, j nodes does not receive input ($n_i = 0$ or $n_j = 0$), node i is shut down $\sigma_i = 0$. This top-down and bottom-up control between different modules is quantified by the following control rule which acts as a logical AND between two controlling nodes (we consider $k_i^{\text{out}} = \{0, 1\}$, see Fig. 1b):

$$\sigma_i = n_i n_j, \quad \text{control rule for } k_i^{\text{out}} = 1. \quad (1)$$

Because not all nodes participate in the control of other nodes, a certain fraction of them (determined by the degree distribution $P(k_i^{\text{out}})$) do not establish intermodular links with other nodes, $k_i^{\text{out}} = 0$. These nodes without control-links (Fig. 1b) activate as soon as they receive an external input, that is

$$\sigma_i = n_i, \quad \text{control rule for } k_i^{\text{out}} = 0. \quad (2)$$

Generalization of the control rule to more than one control link per node can be done in different ways. Here, we consider that a node is activated ($\sigma_i = 1$) iff it receives the input $n_i = 1$ and at least one among the nodes j in another module connected to i via a control link receives also an input, i.e. $n_j = 1$. Otherwise i is not activated (Fig. 1b). Mathematically:

$$\sigma_i = n_i \left[1 - \prod_{j \in \mathcal{F}(i)} (1 - n_j) \right], \quad \text{general control rule (3)}$$

where $\mathcal{F}(i)$ is the set of k_i^{out} nodes connected to i via intermodular control links. In the following, we always refer to the general control model Eq. (3), unless stated otherwise.

The distinction between n_i and σ_i models the initial sensory inputs (n_i), and the final state response of the brain (σ_i) to those stimuli from top-down and bottom-up influences [6]. Thus, the final state of the brain network σ_i encodes the brain's interpretation of the world by modulating external input n_i via controls Eq. (3) from other cortical areas (Fig. 1c). We note that a general model should explain brain activation even when no external input is applied to the NoN (e.g. in resting state brain). This may be accounted for by a dynamical system that drives the NoN into stable attractors, which in resting state may no need external input anymore.

Apart from receiving inputs n_i and controlling other nodes via Eq. (3), active nodes can also broadcast information to the network. When all nodes are active, the information sent by a node can reach the whole brain NoN. If some nodes become inactive, i.e. $\sigma_i = 1 \rightarrow \sigma_i = 0$, the remaining active nodes group together in disjoint components of active nodes, such that information starting from an active node in one active component cannot reach another active node in a different active component. We quantify the global communication efficiency of the brain NoN with the size of the largest (giant) mutually-connected active-component G made of active nodes $\sigma_i = 1$ (Stage 3 in Figs. 1c) [9–11]. By strict definition, G could be (almost) the entire brain, e.g., a visual stimulus sets off emotional cues, memory areas, etc. In what follows, we will restrict the NoN to specific systems of interest in the brain, like the visual or motor system, which are identifiable by fMRI methods for a particular single task.

Each configuration of active/inactive nodes $\vec{\sigma} = (\sigma_1, \dots, \sigma_N)$ is associated to a specific working mode of the brain. The plethora of different functions dynamically executed by the brain [4–7] results in the moment-by-moment changes of the configuration $(\sigma_1, \dots, \sigma_N)$, and thus in different values of G . The crux of the matter is that, for typical input configurations $\vec{n} = (n_1, \dots, n_N)$ —i.e., the ones produced by the majority of the external (e.g. visual) inputs— G has to be large enough for a global integration of information from distributed areas in the brain. In other words, the brain NoN has to remain globally activated during the acquisition of different inputs, meaning that G has to be robust, and the more robust the more states the brain can achieve. Therefore, a model of a brain NoN must be able to capture such robustness.

In our statistical mechanics approach, being robust means that the brain should develop an extensive G for typically sampled configurations of the external inputs. As a first approximation, we assume that these inputs are sampled from a flat (random) distribution of \vec{n} . Thus, we first study the robustness of the NoN across the configurations of states typically sampled by the brain. The problem then becomes a classical percolation study of the NoN [10] following the activation/inactivation rule of Eq. (3). Having established our model in the normal brain under typical inputs, we will then move to disease states, which impede global communication by annihilating focal essential areas in G [13, 14].

We calculate G induced by typical random configurations of inputs \vec{n} as a function of the fraction $q = 1 - \langle n \rangle$ of zero inputs (these zero inputs are analogous to removed nodes in classical percolation [9–11]) and we show that G remains sizeable even for high values of q , thus probing the robustness of the model NoN. At a critical value q_{rand} , we find the random percolation critical point $G(q_{\text{rand}}) = 0$ [9–11] separating a globally connected phase with non-zero $G(q < q_{\text{rand}}) > 0$ from a disconnected phase $G(q > q_{\text{rand}}) = 0$ composed of fragmented

sub-extensive clusters with no giant component in the thermodynamic limit. The most robust NoN is tantamount to a system with no disconnected phase, i.e., with a large value of q_{rand} , ideally $q_{\text{rand}} = 1$. That is, the brain is robust if it can sustain a well-defined giant connected component for as many typical inputs as possible.

The dynamics of information flow in the NoN is defined as follows. Generally speaking, each node processes activity from neighboring nodes. Here, we abstract this coding process by considering that nodes receive information from other nodes via “messages” containing the information about their membership in G . Based on the information they receive, nodes broadcast further messages, until they eventually agree on who belongs to G across the whole network. Since there are two types of links, we define two types of messages: $\rho_{i \rightarrow j} \in \{0, 1\}$ running along an intramodular link, and $\varphi_{i \rightarrow j} \in \{0, 1\}$ running along an intermodular control link, where $\{0, 1\}$ represents a {no, yes} “*I belong to G*” message, respectively (Fig. 1a).

In this view, the existence of an extensive giant mutually-connected component across the NoN, $G > 0$, expresses a percolation phase produced by the binding of activation patterns across different modules in a distributed emergent global system. Under this interpretation, perception is not the responsibility of any particular cortical area but is an emergent critical property of the percolation of memberships interchanged across all members of G [15]. The percolation critical point q_{rand} can be interpreted as the transition between a phase of global perception $G > 0$ for $q < q_{\text{rand}}$ and a null perception phase characterized by non-extensive disconnected components and the concomitant $G = 0$ for $q > q_{\text{rand}}$.

The equations governing the information flow in the brain NoN follow the updating rules of the membership messages according to (analytical details in SI Text):

$$\begin{aligned} \rho_{i \rightarrow j} &= \sigma_i \left[1 - \prod_{k \in \mathcal{S}(i) \setminus j} (1 - \rho_{k \rightarrow i}) \prod_{l \in \mathcal{F}(i)} (1 - \varphi_{l \rightarrow i}) \right], \\ \varphi_{i \rightarrow j} &= \sigma_i \left[1 - \prod_{k \in \mathcal{S}(i)} (1 - \rho_{k \rightarrow i}) \prod_{l \in \mathcal{F}(i) \setminus j} (1 - \varphi_{l \rightarrow i}) \right], \end{aligned} \quad (4)$$

where $\mathcal{S}(i) \setminus j$ is the set of $k_i^{\text{in}} - 1$ neighbors of node i in the same module, except j . Equations (4) indicate, for instance, that a positive membership message $\rho_{i \rightarrow j} = 1$ is transmitted from node $i \rightarrow$ node j in the same module (analogously, $\varphi_{i \rightarrow j}$ transmits messages to the other module) if node i is active $\sigma_i = 1$ and if it receives at least one positive message from either a node k in the same module $\rho_{k \rightarrow i} = 1$ or a node l in the other module $\varphi_{l \rightarrow i} = 1$. The logical OR is important; it is the basis for such a robust R-NoN brain model of activation as elaborated below.

To compute G , it is sufficient to know for each node i whether or not it is a member of G , which is encoded in the quantity $\rho_i \in \{0, 1\}$ representing the probability to

belong to $G = \langle \rho_i \rangle$:

$$\rho_i = \sigma_i \left[1 - \prod_{k \in \mathcal{S}(i)} (1 - \rho_{k \rightarrow i}) \prod_{l \in \mathcal{F}(i)} (1 - \varphi_{l \rightarrow i}) \right]. \quad (5)$$

Here we arrive to an important point (illustrated in Fig. 1c), which ultimately explains the robustness of our brain NoN: in our model a node can be active ($\sigma_i = 1$) even if it does not belong to the giant mutually-connected active component G , thus preventing catastrophic cascading effects. This feature of the brain model is supported by neuro-anatomical correlates: the brain responds reasonably well to injuries, for instance, to areas such as the arcuate fasciculus (the white matter tract that connects the two most important language areas – Broca’s and Wernicke’s area). This property is the main difference between our model and previous NoN models [10] describing catastrophic collapse in power-grids [16], as discussed next.

Universality Classes of NoN.— In the model of Ref. [10], a node can be active only if it belongs to the giant component in its own network. Thus, in this model the active/inactive state of a node is controlled by the whole global giant component ρ_i , rather than the local state variable σ_i , Eq. (3), as in our model. This means that in Ref. [10], the state of a node is actually controlled by the whole network [i.e., intermodular controls (therein called dependencies) carry the weight of the extensive giant component]. Analogously, the NoN cannot be built from the $G = 0$ phase, since it would require the existence of extensive components for each network. For this reason, the resulting NoN [10] is fragile; a single inactivation of a node can lead to catastrophic collapse of the whole active giant component (which, we note, can be avoided by strong correlations between the hubs of different networks [9]). Conversely, the model of Eq. (3) allows nodes to be active even if they do not belong to G , i. e., when they belong to non-extensive disconnected components. These small components become crucial to build the $G > 0$ phase from the $G = 0$ phase by adding interlinks to non-extensive components.

Indeed, the model of [10] was proposed to capture the fragility of certain man-made infrastructures, such as the catastrophic collapse of power grids— e.g., the US North-east blackout of 2003 which allegedly started in a single power-line failure as modeled in [10]. The equation to compute G in this catastrophic C-NoN model reads:

$$\rho_i = \sigma_i \left[1 - \prod_{k \in \mathcal{S}(i)} (1 - \rho_{k \rightarrow i}) \right] \left[1 - \prod_{k \in \mathcal{F}(i)} (1 - \varphi_{k \rightarrow i}) \right]. \quad (6)$$

We note that Eq. (6) differs from R-NoN Eq. (5) in that the logical OR has been replaced by the logical AND for message passing in C-NoN.

A third possible model for NoN is the modular model [2] mentioned in the introduction which assumes no difference between intralinks and interlinks as studied in [17]. In this model there are no control-links, therefore, nodes cannot control each other, and the state equals

the input: $\sigma_i = n_i$. This model is described using only the intralink messages, $\rho_{i \rightarrow j}$, corresponding to a single network structure, albeit with modularity [2], and ρ_i is simple given by (no special messages between modules):

$$\rho_i = n_i \left[1 - \prod_{k \in \mathcal{S}(i)} (1 - \rho_{k \rightarrow i}) \right]. \quad (7)$$

We thus arrive to three different universality classes of NoN— R-NoN, C-NoN and modular single network— according to the three models given by Eqs. (5), (6) and (7), respectively, which are defined according to which variable controls the state of node i (σ_i , ρ_i , n_i), see Table I. Among the three universality classes, only R-NoN is robust with the functionality of control across modules via top-down and bottom-up influences.

Robustness of the brain NoN to typical inputs.— We compute $G(q)$ from Eq. (5) when we present different typical random inputs n_i and show that the obtained percolation threshold q_{rand} is close to 1. The results are first tested on synthetic NoN made of Erdős-Rényi (ER) and scale-free (SF) random graphs [1].

Results in Fig. 2a show that our model indeed defines a robust R-NoN characterized by large q_{rand} . Additionally, Fig. 2b compares model R-NoN Eq. (5) with the catastrophic C-NoN universality class Eq. (6) showing that these two models capture two different phenomena, the former robust with larger q_{rand} and second-order phase transitions, the latter catastrophic with smaller q_{rand} with first-order abrupt transitions.

Response to rare events. Neural Collective Influencers.— Having investigated the behavior of the model under typical inputs, we now study the response of the brain NoN to rare events targeting a set of neural collective influencers (NCI). These are rare inputs: an optimal (minimal) set of nodes that when they are shut-down ($n_i = 0$) disintegrates the giant component to $G = 0$ employing the smallest possible fraction of nodes, q_{infl} . This is the process of optimal percolation (rather than classical random percolation treated above) as defined in [18]. The malfunction of these neural influencers could be associated with pathological states of the brain arising from interruption of global communication in the network structure such as depression or Alzheimer’s disease. The underlying conjecture is that these influencers could be responsible for neurological disorders [13, 14].

Universality Class	State Control	Robust	Control functionality
Brain Robust			
R-NoN Eq. (5)	σ_i	YES	YES
Power-Grid Catastrophic			
C-NoN Eq. (6)	ρ_i	NO	YES
Modular Single			
Network Eq. (7)	n_i	YES	NO

TABLE I. Universality classes of NoN.

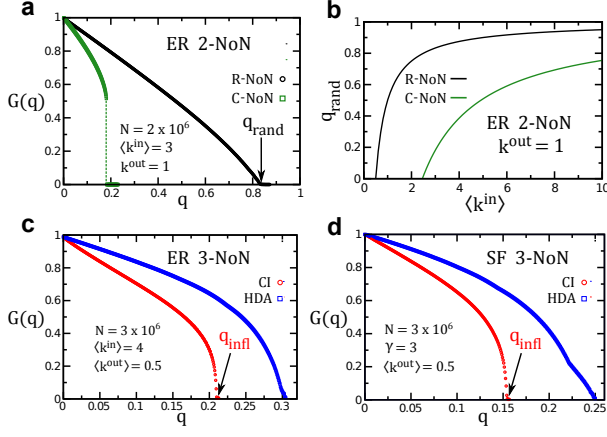


FIG. 2. Robustness and NCI in NoN. a, Robustness of NoN under typical random inputs. Size of the largest active component $G(q)$ for typically sampled inputs \vec{n} for ER 2-NoN (meaning a NoN made of 2 ER networks) for the R-NoN and C-NoN universality classes ($k^{\text{out}} = 1$ for all nodes, one-to-one control links, total size $N = 2 \times 10^6$). The large value of q_{rand} in R-NoN compared to C-NoN confirms the robustness of the former. The transition separating the phases $G = 0$ and $G > 0$ is 2nd-order in R-NoN and 1st-order in C-NoN, reinforcing the fundamental difference (robust *vs* fragile) of these two universality classes (errors are s.e.m. over 10 realizations). **b, Phase diagram for R-NoN and C-NoN.** Behavior of q_{rand} as a function of the average $\langle k^{\text{in}} \rangle$ for the ER 2-NoN in **a**, where each node has $k^{\text{out}} = 1$. Here q_{rand} is the fraction of nodes with zero inputs in one network (nodes in the other network have all nonzero inputs). The difference in q_{rand} between R-NoN and C-NoN ranges from 20% for $\langle k^{\text{in}} \rangle = 10$ to 80% for $\langle k^{\text{in}} \rangle \sim 2.5$. Analytically, we find for R-NoN with $k^{\text{out}} = 1$, $q_{\text{rand}} = 1 - 1/(2\langle k^{\text{in}} \rangle)$. **c, Rare inputs and NCI in ER 3-NoN.** Size of $G(q)$ as a function of the untargeted ($n_i = 0$) nodes q for a NoN of 3 ER networks (total size $N = 3 \times 10^6$). Each network has 10^6 nodes, $\langle k^{\text{in}} \rangle = 4.0$ and $\langle k^{\text{out}} \rangle = 0.5$. We show the CI optimization (red circles, $\ell = 4$) and the high-degree adaptive (HDA) heuristic (blue squares, removal by highest k^{in}) [21]. The arrow marks the position of the minimal fraction of influencers q_{infl} , which is smaller than the HDA centrality (errors are s.e.m. over 10 realizations). Other heuristic centralities perform worse than HDA. **d, Rare inputs and NCI in SF 3-NoN.** $G(q)$ for a NoN with 3 SF networks (total size $N = 3 \times 10^6$). Each network is SF with 10^6 nodes, minimum and maximum degree $k_{\text{min}}^{\text{in}} = 2$ and $k_{\text{max}}^{\text{in}} = 10^3$, and power-law exponent $\gamma = 3$. The node out-degree is Poisson-distributed with average $\langle k^{\text{out}} \rangle = 0.5$ (errors are s.e.m. over 10 realizations). The difference between CI ($\ell = 3$) and HDA is shown; HDA fails to identify 40% of influencers.

At the same time, activating this minimal set of neural influencers, ($n_i = 1$, $\sigma_i = 1$) would optimally broadcast the information to the entire network [19]. Thus, these neural influencers are also the minimal set of nodes that provide integration of global activity in the NoN [15].

Finding this minimal set is a NP-hard combinatorial optimization problem [19]. Here, we follow [18] which

developed the theory of optimal percolation for a system with a single network and solve the problem in a NoN. As opposed to random percolation that identifies the minimal fraction of influencers q_{infl} that, if removed, optimally fragment the giant connected component, i.e., with minimal removals ($n_i = 0$). We note that these neural influencers are statistically rare, i.e., they cannot be obtained by random sampling \vec{n} .

The mapping to optimal percolation [18] allows us to find brain influencers under the approximation of a sparse graph by minimizing the largest eigenvalue $\lambda(q, \vec{n})$ of a modified non-backtracking (NB) matrix [20] $\mathcal{M}_{\rho\varphi} \equiv (\partial \rho_{i \rightarrow j} / \partial \varphi_{k \rightarrow \ell})_{\rho=\varphi=0}$ of the NoN over all configurations of inputs \vec{n} having a fraction q of zero inputs (analytical details in SI Text). The NB matrix $\hat{\mathcal{M}}$ controls the stability of the solution of the broken phase $G = 0$. This solution becomes unstable (i.e., G becomes nonzero) when the largest eigenvalue is 1. The minimal set of influencers \vec{n}_{infl} and their fraction q_{infl} are then found by solving: $\lambda(q_{\text{infl}}, \vec{n}_{\text{infl}}) = \min_{\vec{n}} \lambda(q_{\text{infl}}, \vec{n}) = 1$.

The eigenvalue $\lambda(\vec{n})$ can be efficiently minimized by progressively removing the input ($n_i = 1 \rightarrow n_i = 0$) from the nodes with the highest Collective Influence index $\text{CI}_\ell(i)$ (detailed derivation in SI Text) given by ($z_i \equiv k_i^{\text{in}} + k_i^{\text{out}} - 1$):

$$\text{CI}_\ell(i) = z_i \sum_{j \in \partial \text{Ball}(i, \ell)} z_j + \sum_{\substack{j \in \mathcal{F}(i) : \\ k_j^{\text{out}} = 1}} z_j \sum_{m \in \partial \text{Ball}(j, \ell)} z_m. \quad (8)$$

The collective influence $\text{CI}_\ell(i)$ of node i is determined by two factors (see Fig. 1d). The first one is a *node-centric* contribution, given by the first term in Eq. (8), where $\text{Ball}(i, \ell)$ is the set of nodes inside a ball of radius $\ell > 0$ (ℓ is the distance of the shortest path between two nodes), centered on node i , and $\partial \text{Ball}(i, \ell)$ its frontier. This ball is grown from the central node i by following both intralinks and interlinks, and thus may invade different networks in the NoN. The second factor is a *node-eccentric* contribution, given by the second term in Eq. (8), where the sum runs over all nodes j connected to i by an interlink which have out-degree equal to one $k_j^{\text{out}} = 1$ (this means that nodes j have no other interlinks except to node i). The contribution of each of these j nodes is given by growing another ball $\text{Ball}(j, \ell)$ around them. This last contribution is absent in the single network case [18], and thus, it is a genuine new feature of the brain NoN.

The NCI are formally defined as the nodes in the minimal set upto q_{infl} . To identify them, we start with all $n_i = 1$ and $\sigma_i = 1$ and we progressively remove one by one the inputs (setting $n_i = 1 \rightarrow n_i = 0$) to the nodes having the largest $\text{CI}_\ell(i)$ value if they are active $\sigma_i = 1$. At each step the $\text{CI}_\ell(i)$ values are recomputed, and the algorithm (described in detail in SI Text) stops when $G = 0$ where the NCI set is identified.

We first test our predictions on influencers using synthetically generated ER-NoN and SF-NoN. Figures 2c and 2d show the optimality (smaller q_{inf}) of our predicted set of influencers in comparison with the high-degree centrality [21], a heuristic commonly used in graph analysis of pathological brain networks [14]. The theory allows us to predict the neural collective influence map (NCI-map) of the brain as explained next.

Neural Collective Influence map of the NoN.—

We apply our model to a paradigmatic case of stimulus driven attention [9, 11, 22]. The experiment consists of a dual visual-auditory task performed by 16 subjects (SI Text). Each subject receives simultaneously a visual stimulus and an auditory pitch, to which the subject has to respond with the right hand if a number was larger than a reference and with the left hand if a tone was of high frequency.

The rationale to choose this experiment, where stimuli are received simultaneously, is that this imposes to select an appropriate response order with consequent deployment of high level control modules in the brain [22]. This effect emphasizes the role of top-down control of intermodular links that is the main effect we are trying to capture in our model.

The brain NoN was inferred from the brain activity recorded through functional magnetic resonance imaging (fMRI). Nodes in the NoN represent fMRI voxels whose size is given by the normalized spatial resolution of the fMRI scan $2 \times 2 \times 2 \text{ mm}^3$. Pairwise cross-correlation between the BOLD signals of two nodes represents only indirect correlations (known as the functional connectivity network) capturing the weighted sum of all possible direct interactions between two nodes that could arise from the underlying unknown structural network and others interactions modulating the activity of neurons [7]. In order to construct the brain NoN we infer the strength of these interactions between nodes by using machine learning maximum-entropy methods [23–25], where we maximize the likelihood of the interactions between nodes given the observed pattern of fMRI cross-correlations (full details in SI Text). The resulting NoN is shown in Figs. 3a and b, which are then used to identify the NCI in the brain network activated for this particular task.

In all subjects we observe (Fig. 3a,b): (a) a network partially covering the anterior cingulate (AC) region, recruited for decision making and therefore processing top-down and bottom-up control; (b) a network covering the medial part of the posterior parietal cortex (PPC) which receives somatosensory inputs and sends the output to areas of the frontal motor cortex to control particular movements of the arms; and (c) a network covering the medial part of the posterior occipital cortex (area V1/V2), along the calcarine fissure, which is responsible for processing visual information at lower input levels (an additional auditory network was also observed, see SI Text).

We apply our theory to the AC-PPC-V1/V2 3NoN to first test the robustness under typical inputs and then

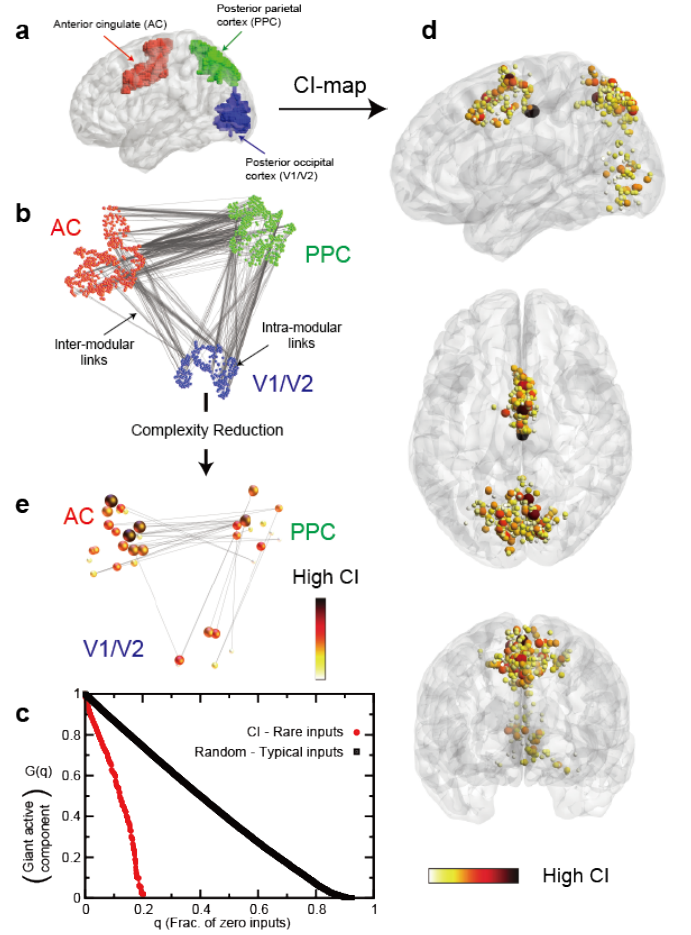


FIG. 3. Brain-NoN. a, 3NoN in dual-task fMRI experiment. Spatial location of the 3 main networks for a typical subject (as opposed to averaging over all subjects as in **d**) showing the anterior cingulate (AC, red), posterior parietal cortex (PPC, green), and posterior occipital visual areas V1/V2 (blue). This 3NoN structure appears consistently for all 16 subjects. Nodes in the NoN represent voxels in the fMRI BOLD signal of normalized size $2 \times 2 \times 2 \text{ mm}^3$. **b, Topology of the 3NoN.** Same as **a**, but in the network representation with interlinks in gray. Number of nodes in NoN is $N = 1,134$, $\langle k^{\text{in}} \rangle = 3.2$, and $\langle k^{\text{out}} \rangle = 2.5$. **c, Robustness and NCI.** Size of the largest active cluster $G(q)$ as a function of the untargeted ($n_i = 0$) nodes q following CI optimization (red curve, $\ell = 3$) and following typical random states (black, random percolation). **d, NCI-map of the human brain** averaged over 16 subjects. The color code ranges from 0 to 5.2 and represents the number of subjects a node appears in the ranked NCI set (see SI Text). High-CI influential regions are located mainly in the AC module for processing top-down control, whereas the influential nodes are rarely located in the lower-level V1/V2 region. The PPC region contains a portion of influential nodes closer to AC. **e, Complexity reduction** to top NCI nodes. Controlling links between different networks are mainly mediated by top influencers.

obtain the NCI (rare inputs). Indeed, the obtained brain 3NoN is very robust to typical inputs as shown by the large (close to one) $q_{\text{rand}} \approx 0.9$ in Fig. 3c, black curve. On the other hand, the theory is able to localize the minimal set of NCI with $q_{\text{infl}} \approx 0.2$, Fig. 3c, red curve. Using these influential nodes we construct the NCI-map averaging over all subjects. The emerging NCI-map averaged over the 16 subjects is portrayed in Fig. 3d (details in SI Text). We find that the main influence region (high CI) is located mainly in the AC module as expected, since AC is a central station of top-down control. The areas of high influence extends also to a portion of the PPC involved in both top-down and bottom-up control, and it is less prominent in the V1/V2 areas, which are mostly involved in processing input information and bottom-up interactions. Therefore, the NCI-map of the brain suggests that control is deployed from the higher level module (AC) towards certain strategic locations in the lower ones (PPC-V1/V2), and these locations can be predicted by network theory. The complexity reduction obtained by coarse-graining the whole NoN to the top NCI in Fig. 3e highlights the predicted strategic areas in the brain.

DISCUSSION

We present a minimal model of a robust NoN to describe the integration of brain modules via control in-

terconnections. The key point of the model is that a node can be active even if it does not belong to the giant mutually-connected active-component so that cascades are not fatal. While our model is expressed *in abstracto* by logic relations, it is able to make falsifiable predictions, e.g., the location of the most influential neural nodes involved in information processing in the brain. If confirmed experimentally, our results may have applications of clinical interest, in that they may help to design therapeutic protocols to handle pathological network conditions and to retune diseased network dynamics in specific neurological disorders with interventions targeted to the activity of the influential nodes predicted by network theory. On the theoretical side, further extensions of our model are also possible. For instance, the model could be enriched by incorporating temporal dependence of brain activation, which are relevant for the theoretical description of synaptic plasticity [26].

Acknowledgment. We thank S. Canals, S. Havlin and L. Parra for discussions and M. Sigman for providing the data. This work was supported by NSF Grants PHY-1305476 and IIS-1515022; NIH-NIBIB Grant 1R01EB022720, NIH-NCI U54CA137788/U54CA132378; and ARL Grant W911NF-09-2-0053 (ARL Network Science CTA). The Boston University work was supported by NSF Grants PHY-1505000, CMMI-1125290, and CHE-1213217, and by DTRA Grant HDTRA1-14-1-0017 and DOE Contract DE-AC07-05Id14517.

-
- [1] Caldarelli G, Vespignani A (2007) *Large Scale Structure and Dynamics of Complex Networks: From Information Technology to Finance and Natural Science* (World Scientific, Singapore).
 - [2] Newman MEJ (2006) Modularity and community structure in networks. *Proc. Natl Acad. Sci. USA* 103: 8577-8582.
 - [3] Tononi G, Sporns O, Edelman GM (1994) A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proc. Natl Acad. Sci. USA* 91: 5033-5037.
 - [4] Treisman A (1996) The binding problem. *Curr. Opin. Neurobiol.* 6: 171-178.
 - [5] Dehaene S, Naccache L (2001) Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition* 79: 1-37.
 - [6] Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3: 201-215.
 - [7] Bullmore E, Sporns O (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* 10: 186-198.
 - [8] Russo R, Herrmann HJ, de Arcangelis L (2014) Brain modularity controls the critical behavior of spontaneous activity. *Sci. Rep.* 4: 4312.
 - [9] Reis SDS, et al. (2014) Avoiding catastrophic failure in correlated network of networks. *Nature Phys.* 10: 762-767.
 - [10] Buldyrev SV, Parshani R, Paul G, Stanley HE, Havlin S (2010) Catastrophic cascade of failures in interdependent networks. *Nature* 464: 1025-1028.
 - [11] Gallos LK, Makse HA, Sigman M (2012) A small world of weak ties provides optimal global integration of self-similar modules in functional brain networks. *Proc. Natl Acad. Sci. USA* 109: 2825-2830.
 - [12] Gilbert CD, Sigman M (2007) Brain states: top-down influences in sensory processing. *Neuron* 54: 677-696.
 - [13] Stam CJ (2014) Modern network science of neurological disorders. *Nat. Rev. Neurosci.* 15: 683-695.
 - [14] van den Heuvel MP, Sporns O (2013) Network hubs in the human brain. *Trends Cogn. Sci.* 17: 683-696.
 - [15] Crick F, Koch C (2003) A framework for consciousness. *Nature Neurosci.* 6 119-126.
 - [16] Rosato V, et al. (2008) Modeling interdependent infrastructures using interacting dynamical models. *Int. J. Critical Inf. Syst.* 4: 63-79.
 - [17] Leicht EA, D'Souza RM (2009) Percolation on interacting networks. arXiv:0907.0894.
 - [18] Morone F, Makse HA (2015) Influence maximization in complex networks through optimal percolation. *Nature* 524: 65-68.
 - [19] Kempe D, Kleinberg J, Tardos E (2003) Maximizing the spread of influence through a social network. *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (Association for*

- Computing Machinery, New York*), p137-143.
- [20] Hashimoto K (1989) Zeta functions of finite graphs and representations of p-adic groups. *Adv. Stud. Pure Math.* 15: 211-280.
 - [21] Albert R, Jeong H, Barabási A-L (2000) Error and attack tolerance of complex networks. *Nature* 406: 378-382.
 - [22] Sigman M, Dehaene S (2008) Brain mechanisms of serial and parallel processing during dual-task performance. *J. Neurosci.* 28: 7585-7598.
 - [23] Schneidman E, Berry MJ, Segev R, Bialek W (2006) Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* 440: 1007-1012.
 - [24] Sarkar S, Chawla S, Xu D (2015) On inferring structural connectivity from brain functional-MRI data. arXiv:1502.06659
 - [25] Robinson PA, Sarkar S, Pandejee GM, Henderson J (2014) Determination of effective brain connectivity from functional connectivity with application to resting state connectivities. *Phys. Rev. E* 90: 012707.
 - [26] Min B, Moreno A, Morone F, Perez-Ramirez U, Perez-Cervera L, Parra LC, Holodny A, Canals S, Makse HA (2017) Finding essential nodes for integration in the brain using network optimization theory (Submitted).

Supplementary Info

MESSAGE PASSING IN THE BRAIN-NO

The classification of connections into intramodule, *intra-links*, and intermodule, *inter-links*, together with an introduction of the mathematical model describing robust brain Network of Networks (NoN) was provided in the main article. In the present section we further expand the explanation of the NoN model and the derivation of the message passing equations describing the information flow in the brain. In the following sections, we then tackle the problem of finding the most influential nodes in the brain-NoN with general configuration of intra- and inter-links. We conclude with an explanation of the numerical tests and the construction of the CI-map of the brain.

In what follows, we consider two modules A and B, interconnected by undirected inter-links, where each module is an independent network made up of N_A respectively N_B nodes connected via intra-links ($N = N_A + N_B$). The theoretical approach and indeed the obtained collective influence formula readily carry over to arbitrary numbers of modules.

Throughout most of the supplementary sections, we adopt the convention to explicitly show the node's belonging to either module, i.e. every index i_A representing a node will be accompanied by the network label to which the node belongs. Moreover, we denote a node's degree of undirected intra-links by $k_{i_A}^{\text{in}}$ and undirected inter-links degree by $k_{i_A}^{\text{out}}$. Furthermore, the input variable $n_{i_A} = 1, 0$ specifies whether node i_A receives an external input ($n_{i_A} = 1$) or not ($n_{i_A} = 0$). It is understood that the same terminology applies equivalently to nodes i_B in module B.

Following the definition of our brain model, we assume that a node i_A which is connected to one or several nodes from the other module is activated ($\sigma_{i_A} = 1$) if it receives an input ($n_{i_A} = 1$) and at least one among the nodes j_B connected to it via an inter-link also receives an input ($n_{j_B} = 1$), as depicted in Fig. 1b. In other words, a node with one or several inter-link dependencies is inactivated when it does not receive the input ($n_{i_A} = 0$), or when the last of its neighbors in the other module ceases to receive an external input. This interaction is mathematically formalized by the concept of the state variable σ_{i_A} :

$$\sigma_{i_A} = n_{i_A} \left[1 - \prod_{j_B \in \mathcal{F}(i_A)} (1 - n_{j_B}) \right], \quad (9)$$

where $\mathcal{F}(i_A)$ denotes the set of nodes in module B connected to i_A via an inter link. For the case that node i_A has exactly one inter-link to one node j_B in module B, the above equation reduces to

$$\sigma_{i_A} = n_{i_A} n_{j_B}, \quad \text{for one-to-one connections.} \quad (10)$$

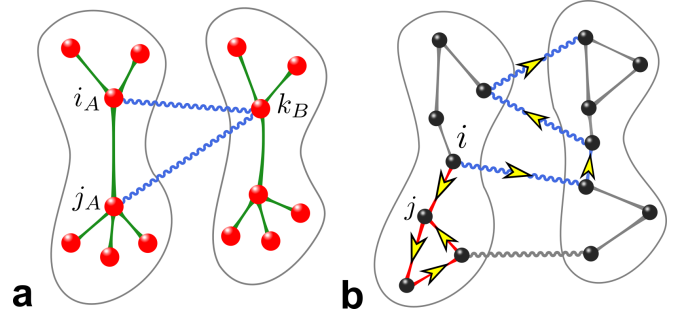


FIG. 4. **a** Simple NoN illustrating the activation rule Eq. (9). **b** Two NB walks of length $\ell = 4$, centered in node i in the 2-NoN. Note that the red walk visits node j twice, hence it contains a NB loop. However, as shown in [18], NB walks with loops can be neglected in the cost energy function to leading order $\mathcal{O}(N)$.

By convention, we also agree to include in the above equation for σ_{i_A} the case where node i_A does not have any inter-links $k_{i_A}^{\text{out}} = 0$. In this case, we simply equate

$$\sigma_{i_A} = n_{i_A}, \quad \text{for } k_{i_A}^{\text{out}} = 0. \quad (11)$$

Alternatively, we can say that products over empty sets $\mathcal{F}(i_A) = \emptyset$ default to zero. This is an important feature of the model, namely that a fraction of nodes determined by $\langle k^{\text{out}} \rangle$ are not involved in control.

In order to get a better understanding of the state variable σ_{i_A} , we consider the following example of the simple NoN depicted in Fig. 4a. For this particular case, we have

$$\begin{aligned} \sigma_{i_A} &= n_{i_A} n_{k_B}, \\ \sigma_{j_A} &= n_{j_A} n_{k_B}, \\ \sigma_{k_B} &= n_{k_B} [1 - (1 - n_{i_A})(1 - n_{j_A})], \end{aligned} \quad (12)$$

and the remaining nodes l with no inter-links, $k_l^{\text{out}} = 0$, have $\sigma_l = n_l$.

As can be seen, when the nodes in A receive input $n_{i_A} = n_{j_A} = 1$ but node k_B does not, $n_{k_B} = 0$, this configuration of external inputs affects all state variables $\sigma_{k_B} = \sigma_{i_A} = \sigma_{j_A} = 0$. On the other hand, keeping $n_{j_A} = n_{k_B} = 1$ and removing the input $n_{i_A} = 0$ only affects the state of node i_A by switching it to inactive $\sigma_{i_A} = 0$ since node k_B is connected to another node in module A, namely j_A , and hence $\sigma_{k_B} = 1$ is active together with $\sigma_{j_A} = 1$.

Let us now turn our attention to the messages, representing information broadcasted between active nodes within the same module or between active nodes in different modules. The distinction between intramodule and intermodule messages naturally arises due to the conceptual difference between intra-links and inter-links and is reflected in the corresponding distinction between messages $\rho_{i_A \rightarrow j_A}$ sent along intra-links and messages $\varphi_{i_A \rightarrow j_B}$ transmitted across inter-links (Fig. 1a).

It is clear that when all nodes are initially active, the information is able to circulate in the entire NoN. On the other hand, as individual nodes are sequentially turned off, the remaining active nodes are progressively fragmented into disconnected clusters and as a result the information can no longer be broadcasted globally. The efficiency to communicate globally can thus be represented by the size of the *largest (giant) connected cluster of active nodes* G across all modules constituting the NoN, as depicted in Figs. 1c, e.

Formally, we denote

$$\begin{aligned} \rho_{i_A \rightarrow j_A} &\equiv \text{probability that } i_A \text{ is connected to } G \\ &\quad \text{other than via in-neighbor } j_A, \\ \varphi_{i_A \rightarrow j_B} &\equiv \text{probability that } i_A \text{ is connected to } G \\ &\quad \text{other than via out-neighbor } j_B. \end{aligned} \quad (13)$$

The size of the mutual giant active component G in turn is entirely determined by the solution of a set of $2M$ self-consistent message passing equations, where M is the total number of intra-links and inter-links in the NoN.

The derivation of the set of message passing equations corresponding to our model is provided next. Let us therefore consider two nodes in the NoN, say i_A and j_A , connected by an intra-link. A node i_A can send information only if it is active, i.e. if $\sigma_{i_A} = 1$, and hence the relative message $\rho_{i_A \rightarrow j_A}$ must be proportional to σ_{i_A} . Now, assuming that node i_A is active, it can send a message to node j_A only if it receives a message by at least one of its intra-link neighbors other than j_A OR one of its inter-links neighbors. Thus, the self-consistent equations describing the information flow in the brain NoN are given by

$$\begin{aligned} \rho_{i_A \rightarrow j_A} &= \sigma_{i_A} \left[1 - \prod_{k_A \in \mathcal{S}(i_A) \setminus j_A} (1 - \rho_{k_A \rightarrow i_A}) \prod_{k_B \in \mathcal{F}(i_A)} (1 - \varphi_{k_B \rightarrow i_A}) \right], \\ \varphi_{i_A \rightarrow j_B} &= \sigma_{i_A} \left[1 - \prod_{k_A \in \mathcal{S}(i_A)} (1 - \rho_{k_A \rightarrow i_A}) \prod_{k_B \in \mathcal{F}(i_A) \setminus j_B} (1 - \varphi_{k_B \rightarrow i_A}) \right], \end{aligned} \quad (14)$$

where $\mathcal{S}(i_A)$ is the set of intra-link neighbors of node i_A and $\mathcal{F}(i_A)$ is the set of node i_A 's inter-links neighbors in module B. The remaining message passing equations can be obtained by interchanging the labels for the modules A and B. We note en passant that products over empty sets $\mathcal{S}(i_A) = \emptyset$ or $\mathcal{F}(i_A) = \emptyset$ in the above message passing equations default to one, due to the underlying logical OR in our model.

The size of the mutual giant component G across all modules of the NoN can then be computed from the fixed point solution for the intra-link and inter-link messages satisfying the above self-consistent message passing equations (14). Explicitly, it is given by

$$G = \left(\sum_{i_A=1}^{N_A} \rho_{i_A} + \sum_{i_B=1}^{N_B} \rho_{i_B} \right) / (N_A + N_B), \quad (15)$$

where the probability $\rho_{i_A} = 0, 1$ for a node i_A to belong to the largest connected active cluster is computed as

$$\rho_{i_A} = \sigma_{i_A} \left[1 - \prod_{k_A \in \mathcal{S}(i_A)} (1 - \rho_{k_A \rightarrow i_A}) \prod_{k_B \in \mathcal{F}(i_A)} (1 - \varphi_{k_B \rightarrow i_A}) \right], \quad (16)$$

which can be obtained from the expression for the intra-link message in Eq. (14) by including the contribution of $\rho_{j_A \rightarrow i_A}$ as well.

Strictly speaking, the above message passing equations are valid only under the assumption that the messages are independent, which is true for locally tree-like networks, including the thermodynamic limit of the class of Erdős-Rényi and scale-free networks as well as the configuration model (the maximally random graphs with a given degree distribution [28]) which contain loops that grow logarithmically in the system size [32]. Nevertheless, it is generally accepted, and confirmed by previous implementations of CI on single networks [18], that results obtained for tree-like graphs apply quite well also for loopy networks [29–31].

Next, we turn our attention to two related, but fundamentally different models. One of them [10], inspired by the power grid [16], can be simply obtained from the message passing equations (14) by replacing the underlying logical OR with a logical AND, as follows

$$\begin{aligned} \rho_{i_A \rightarrow j_A} &= \sigma_{i_A} \left[1 - \prod_{k_A \in \mathcal{S}(i_A) \setminus j_A} (1 - \rho_{k_A \rightarrow i_A}) \right] \left[1 - \prod_{k_B \in \mathcal{F}(i_A)} (1 - \varphi_{k_B \rightarrow i_A}) \right], \\ \varphi_{i_A \rightarrow j_B} &= \sigma_{i_A} \left[1 - \prod_{k_A \in \mathcal{S}(i_A)} (1 - \rho_{k_A \rightarrow i_A}) \right] \left[1 - \prod_{k_B \in \mathcal{F}(i_A) \setminus j_B} (1 - \varphi_{k_B \rightarrow i_A}) \right]. \end{aligned} \quad (17)$$

In this model, an active node i_A with inter-links to the other module can send a message $\rho_{i_A \rightarrow j_A}$ to node j_A only if it receives a message by at least one of its intra-link neighbors other than j_A AND one of its inter-link neighbors.

Similarly, the probability ρ_{i_A} for a node i_A to belong to the giant mutually connected active component G can for this model [10] be obtained by replacing the inherent logical OR in Eq. (16) with the connective AND:

$$\rho_{i_A} = \sigma_{i_A} \left[1 - \prod_{k_A \in \mathcal{S}(i_A)} (1 - \rho_{k_A \rightarrow i_A}) \right] \left[1 - \prod_{k_B \in \mathcal{F}(i_A)} (1 - \varphi_{k_B \rightarrow i_A}) \right]. \quad (18)$$

We emphasize that Eqs. (17) and (18) are generalizations of the model [10], which considers only one-to-one inter-link (therein called dependencies), to arbitrary numbers of inter-links.

The third candidate for a NoN to be considered is the simplest possible model, which assumes no difference between intramodule and intermodule connections [2, 17] and hence it can be described using only the intra-link messages $\rho_{i \rightarrow j}$, which in this case run along links both within and across modules. Moreover, since there are no dependency links in this model and nodes do not control each other, the state of a node simply equals its input

$\sigma_i = n_i$. The corresponding message passing equations read

$$\rho_{i \rightarrow j} = n_i \left[1 - \prod_{k \in \mathcal{S}(i) \setminus j} (1 - \rho_{k \rightarrow i}) \right], \quad (19)$$

where for simplicity we dropped the unneeded distinction between different module labels.

The probability ρ_i for a node i to belong to the giant mutually connected active cluster G can again be obtained by taking into account also the contribution from $\rho_{j \rightarrow i}$, as in

$$\rho_i = n_i \left[1 - \prod_{k \in \mathcal{S}(i)} (1 - \rho_{k \rightarrow i}) \right]. \quad (20)$$

We conclude this discussion by pointing out that the message passing approach presented in this section not only allows to study percolation in NoN in a simple and compact way, but it also allows to treat the non-random removal of inputs and hence investigate the effect of atypical or rare configurations of inputs on the brain state. Moreover, the message passing approach allows for an intuitive interpretation in terms of information flow and can be easily adapted to include changes in the model as well.

Finally, we recall that the size of the giant mutually connected active component and indeed the NoN's global communication efficiency is a function of the input variables n_{i_A} of each node comprising the NoN. The aim of the next section is thus to find and rank the minimal set of nodes whose disruption ($n_{i_A} = 1 \rightarrow n_{i_A} = 0$) leads to a breakdown of the NoN's global communication capacity in the most efficient way. We call such nodes *influencers*.

THEORY OF COLLECTIVE INFLUENCE IN THE BRAIN-NON

Derivation of the cost energy function of influence

Finding the minimal set of influencers, whose inactivation results in a breakdown of the NoN's global communication efficiency, is a NP-hard combinatorial optimization problem originally posed by Kempe *et al.* [19] in the context of maximization of influence in social network, that is very difficult to solve in general. In particular, direct minimization of the size of the mutual giant component over the configurations of inputs $\vec{n} = \{n_{1_A}, \dots, n_{N_A}, n_{1_B}, \dots, n_{N_B}\}$ is untractable, since an explicit functional form of $G(\vec{n})$ is not feasible.

Instead, the problem of identifying the set of influencers in the brain NoN can be mapped onto the problem of optimal percolation [18], which, in turn, can be solved by minimizing the largest eigenvalue $\lambda(\vec{n})$ of the non-backtracking (NB) matrix of the NoN [18]. The NB matrix controls the stability of the broken solution $G = 0$ which corresponds to

$\{\rho_{i_A \rightarrow j_A}\} = \{\rho_{i_B \rightarrow j_B}\} = \{\varphi_{i_A \rightarrow j_B}\} = \{\varphi_{i_B \rightarrow j_A}\} = 0$ and is defined by taking partial derivatives in the message passing equations (14), as follows:

$$\hat{\mathcal{M}} \equiv \begin{pmatrix} \frac{\partial \rho_{k_A \rightarrow l_A}}{\partial \rho_{i_A \rightarrow j_A}} & \frac{\partial \rho_{k_B \rightarrow l_B}}{\partial \rho_{i_A \rightarrow j_A}} & \frac{\partial \varphi_{k_A \rightarrow l_B}}{\partial \rho_{i_A \rightarrow j_A}} & \frac{\partial \varphi_{k_B \rightarrow l_A}}{\partial \rho_{i_A \rightarrow j_A}} \\ \frac{\partial \rho_{k_A \rightarrow l_A}}{\partial \rho_{i_B \rightarrow j_B}} & \frac{\partial \rho_{k_B \rightarrow l_B}}{\partial \rho_{i_B \rightarrow j_B}} & \frac{\partial \varphi_{k_A \rightarrow l_B}}{\partial \rho_{i_B \rightarrow j_B}} & \frac{\partial \varphi_{k_B \rightarrow l_A}}{\partial \rho_{i_B \rightarrow j_B}} \\ \frac{\partial \rho_{k_A \rightarrow l_A}}{\partial \varphi_{i_A \rightarrow j_B}} & \frac{\partial \rho_{k_B \rightarrow l_B}}{\partial \varphi_{i_A \rightarrow j_B}} & \frac{\partial \varphi_{k_A \rightarrow l_B}}{\partial \varphi_{i_A \rightarrow j_B}} & \frac{\partial \varphi_{k_B \rightarrow l_A}}{\partial \varphi_{i_A \rightarrow j_B}} \\ \frac{\partial \rho_{k_A \rightarrow l_A}}{\partial \varphi_{i_B \rightarrow j_A}} & \frac{\partial \rho_{k_B \rightarrow l_B}}{\partial \varphi_{i_B \rightarrow j_A}} & \frac{\partial \varphi_{k_A \rightarrow l_B}}{\partial \varphi_{i_B \rightarrow j_A}} & \frac{\partial \varphi_{k_B \rightarrow l_A}}{\partial \varphi_{i_B \rightarrow j_A}} \end{pmatrix} \bigg|_{G=0} \quad (21)$$

We note that the NB matrix $\hat{\mathcal{M}}_{i \rightarrow j, k \rightarrow l}$ is defined over the space of links (see below) and has non-zero entries only when $(i \rightarrow j, k \rightarrow l)$ form a pair of consecutive non-backtracking edges, i.e. $(i \rightarrow j, j \rightarrow l)$ with $i \neq l$ [18] (see also Fig. 4b). Moreover, powers of the NB matrix count the number of non-backtracking walks of a given length much in the same way as powers of adjacency matrices count the number of paths.

The minimization of $\lambda(\vec{n})$ is performed over the space of input configurations \vec{n} satisfying the condition $(\sum_{i_A} n_{i_A} + \sum_{i_B} n_{i_B}) / (N_A + N_B) = 1 - q$, where q denotes the fraction of zero inputs. The zero solution of the message passing equations, corresponding to a particular configuration \vec{n} , is stable if the largest eigenvalue of the respective NB matrix satisfies $\lambda(\vec{n}) < 1$. Therefore, the optimal configuration \vec{n}_{infl} of influencers (for which $n_{i_A}, n_{j_B} = 0$), can be found by solving

$$\lambda(q_{\text{infl}}, \vec{n}_{\text{infl}}) \equiv \min_{\vec{n}} \lambda(q_{\text{infl}}, \vec{n}) = 1, \quad (22)$$

where q_{infl} denotes the minimal fraction of zero inputs, i.e. the influencers. To keep notation light, we shall from now on omit q in $\lambda(q, \vec{n}) \equiv \lambda(\vec{n})$, which we assume to be kept fixed.

In order to arrive at an explicit expression for the largest eigenvalue, we observe that $\lambda(\vec{n})$ determines the growth rate of an arbitrary non-zero vector \vec{w}_0 after ℓ iterations with the NB matrix $\hat{\mathcal{M}}$, provided it has non-vanishing projection onto the corresponding eigenvector. More precisely, the following equality holds according to the Power Method:

$$\lambda(\vec{n}) = \lim_{\ell \rightarrow \infty} \left[\frac{\langle \mathbf{w}_0 | \hat{\mathcal{M}}^\ell | \mathbf{w}_0 \rangle}{\langle \mathbf{w}_0 | \mathbf{w}_0 \rangle} \right]^{1/\ell}, \quad (23)$$

where $|\mathbf{w}_0\rangle = \vec{w}_0$ denotes the usual column vector and $\langle \mathbf{w}_0| = \vec{w}_0^T$ denotes the corresponding row vector.

For finite ℓ we define $\langle \mathbf{w}_0 | \hat{\mathcal{M}}^\ell | \mathbf{w}_0 \rangle$ to be the cost energy function of influence at order- ℓ and denote the ℓ -dependent approximation to the largest eigenvalue

$$\lambda_\ell(\vec{n}) \equiv \left[\frac{\langle \mathbf{w}_0 | \hat{\mathcal{M}}^\ell | \mathbf{w}_0 \rangle}{\langle \mathbf{w}_0 | \mathbf{w}_0 \rangle} \right]^{1/\ell}. \quad (24)$$

In order to derive an analytical expression for $\lambda_\ell(\vec{n})$, it is convenient to elevate the NB matrix $\hat{\mathcal{M}}$ from the above implicit representation over

the space of $2(M_A+M_B+M_{AB}) \times 2(M_A+M_B+M_{AB})$ links, where M_A , M_B and M_{AB} respectively denote the number of intramodule and intermodule links, and embed it into an enlarged space of dimension $(N_A+N_B) \times (N_A+N_B) \times (N_A+N_B) \times (N_A+N_B)$ [18].

In this enlarged space, the non-vanishing blocks corresponding to the NB matrix of our NoN are obtained from Eqs. (14) and are given by (the remaining blocks can be obtained by interchanging the module labels)

$$\begin{aligned} \left. \frac{\partial \rho_{k_A \rightarrow l_A}}{\partial \rho_{i_A \rightarrow j_A}} \right|_{G=0} &= \sigma_{k_A} A_{i_A j_A}^{\text{in}} A_{k_A l_A}^{\text{in}} \delta_{j_A k_A} (1 - \delta_{i_A l_A}) \\ \left. \frac{\partial \rho_{k_A \rightarrow l_A}}{\partial \varphi_{i_B \rightarrow j_A}} \right|_{G=0} &= \sigma_{k_A} A_{i_B j_A}^{\text{out}} A_{k_A l_A}^{\text{in}} \delta_{j_A k_A} \\ \left. \frac{\partial \varphi_{k_A \rightarrow l_B}}{\partial \rho_{i_A \rightarrow j_A}} \right|_{G=0} &= \sigma_{k_A} A_{i_A j_A}^{\text{in}} A_{k_A l_B}^{\text{out}} \delta_{j_A k_A} \\ \left. \frac{\partial \varphi_{k_A \rightarrow l_B}}{\partial \varphi_{i_B \rightarrow j_A}} \right|_{G=0} &= \sigma_{k_A} A_{i_B j_A}^{\text{out}} A_{k_A l_B}^{\text{out}} \delta_{j_A k_A} (1 - \delta_{i_B l_B}), \end{aligned} \quad (25)$$

In the above equations A stands for adjacency matrix and the superscript 'in' means that both nodes, represented by the subscript indices, are within the same module, whereas 'out' indicates that they are located in distinct modules. We remind ourselves that the matrix entries at positions (i_A, j_B) and (j_B, i_A) are $A_{i_A j_B}^{\text{out}} = A_{j_B i_A}^{\text{out}} = 1$ if there exists a connection (in this case an inter-link) between nodes i_A and j_B and $A_{i_A j_B}^{\text{out}} = A_{j_B i_A}^{\text{out}} = 0$ if there is no connection between these nodes. The Kronecker deltas reflect the non-backtracking property underlying the message passing equations (14), which essentially arises due to the fact that a message is computed on the basis of incoming messages other than from the destination it is sent to.

Similarly, the intrinsically $2(M_A+M_B+M_{AB})$ dimensional starting vector \vec{w}_0 , can be embedded into a larger space of dimension $(N_A+N_B) \times (N_A+N_B)$. Without loss of generality, we choose $|\mathbf{w}_0\rangle = |\mathbf{1}\rangle$ as starting vector in the Power Method Iteration, which translates to $|\mathbf{w}_0\rangle_{i,j} \equiv (A_{i_A j_A}^{\text{in}}, A_{i_B j_B}^{\text{in}}, A_{i_A j_B}^{\text{out}}, A_{i_B j_A}^{\text{out}})^T$ over the enlarged vector space.

In what follows, we are going to develop the general ℓ -th order expression for the cost energy function of influence corresponding to the NB matrix of our NoN, which reads

$$\hat{\mathcal{M}} = \left(\begin{array}{cccc} \frac{\partial \rho_{k_A \rightarrow l_A}}{\partial \rho_{i_A \rightarrow j_A}} & 0 & \frac{\partial \varphi_{k_A \rightarrow l_B}}{\partial \rho_{i_A \rightarrow j_A}} & 0 \\ 0 & \frac{\partial \rho_{k_B \rightarrow l_B}}{\partial \rho_{i_B \rightarrow j_B}} & 0 & \frac{\partial \varphi_{k_B \rightarrow l_A}}{\partial \rho_{i_B \rightarrow j_B}} \\ 0 & \frac{\partial \rho_{k_B \rightarrow l_B}}{\partial \varphi_{i_A \rightarrow j_B}} & 0 & \frac{\partial \varphi_{k_B \rightarrow l_A}}{\partial \varphi_{i_A \rightarrow j_B}} \\ \frac{\partial \rho_{k_A \rightarrow l_A}}{\partial \varphi_{i_B \rightarrow j_A}} & 0 & \frac{\partial \varphi_{k_A \rightarrow l_B}}{\partial \varphi_{i_B \rightarrow j_A}} & 0 \end{array} \right) \bigg|_{G=0} \quad (26)$$

To this end, we investigate order by order the cost energy function until the general expression becomes ev-

ident. To order $\ell = 1$, we find

$$\begin{aligned} \langle \mathbf{w}_0 | \hat{\mathcal{M}} | \mathbf{w}_0 \rangle &= \sum_{i,j,k,l}^{N_A+N_B} i j \langle \mathbf{w}_0 | \hat{\mathcal{M}}_{i j k l} | \mathbf{w}_0 \rangle_{k l} \\ &= \sum_{i_A}^{N_A} \sum_{j_A}^{N_A} \left[\sum_{k_A}^{N_A} \sum_{l_A}^{N_A} A_{i_A j_A}^{\text{in}} \left(\frac{\partial \rho_{k_A \rightarrow l_A}}{\partial \rho_{i_A \rightarrow j_A}} \right) A_{k_A l_A}^{\text{in}} \right. \\ &\quad \left. + \sum_{k_A}^{N_A} \sum_{l_B}^{N_B} A_{i_A j_A}^{\text{in}} \left(\frac{\partial \varphi_{k_A \rightarrow l_B}}{\partial \rho_{i_A \rightarrow j_A}} \right) A_{k_A l_B}^{\text{out}} \right] \\ &\quad + \sum_{i_B}^{N_B} \sum_{j_A}^{N_A} \left[\sum_{k_A}^{N_A} \sum_{l_A}^{N_A} A_{i_B j_A}^{\text{out}} \left(\frac{\partial \rho_{k_A \rightarrow l_A}}{\partial \varphi_{i_B \rightarrow j_A}} \right) A_{k_A l_A}^{\text{in}} \right. \\ &\quad \left. + \sum_{k_A}^{N_A} \sum_{l_B}^{N_B} A_{i_B j_A}^{\text{out}} \left(\frac{\partial \varphi_{k_A \rightarrow l_B}}{\partial \varphi_{i_B \rightarrow j_A}} \right) A_{k_A l_B}^{\text{out}} \right] \\ &\quad + \{A \leftrightarrow B\}, \end{aligned} \quad (27)$$

where $\{A \leftrightarrow B\}$ means "the same terms as above but with interchanged module labels".

Inserting the relations for the partial derivatives given by Eq. (25) and summing over all independent indices, we obtain the following expression for the cost energy function to lowest order,

$$\begin{aligned} \langle \mathbf{w}_0 | \hat{\mathcal{M}} | \mathbf{w}_0 \rangle &= \sum_{k_A} \sigma_{k_A} (k_{k_A}^{\text{in}} + k_{k_A}^{\text{out}} - 1) k_{k_A}^{\text{in}} + \sigma_{k_A} (k_{k_A}^{\text{in}} + k_{k_A}^{\text{out}} - 1) k_{k_A}^{\text{out}} \\ &\quad + \sum_{k_B} \sigma_{k_B} (k_{k_B}^{\text{in}} + k_{k_B}^{\text{out}} - 1) k_{k_B}^{\text{in}} + \sigma_{k_B} (k_{k_B}^{\text{in}} + k_{k_B}^{\text{out}} - 1) k_{k_B}^{\text{out}}. \end{aligned} \quad (28)$$

At this point, it is worth introducing the following notation, which will appear frequently in subsequent expressions for higher order terms

$$z_{i_A} \equiv (k_{i_A}^{\text{in}} + k_{i_A}^{\text{out}} - 1). \quad (29)$$

This allows us to rewrite even more compactly the final expression for the cost energy function of influence at order $\ell = 1$,

$$\begin{aligned} \langle \mathbf{w}_0 | \hat{\mathcal{M}} | \mathbf{w}_0 \rangle &= \sum_{k_A} \sigma_{k_A} z_{k_A} (k_{k_A}^{\text{in}} + k_{k_A}^{\text{out}}) + \sum_{k_B} \sigma_{k_B} z_{k_B} (k_{k_B}^{\text{in}} + k_{k_B}^{\text{out}}). \end{aligned} \quad (30)$$

We proceed to compute the cost energy function to second order from the square of our NB matrix as follows

$$\langle \mathbf{w}_0 | \hat{\mathcal{M}}^2 | \mathbf{w}_0 \rangle = \sum_{i,j,k,l,m,n}^{N_A+N_B} i j \langle \mathbf{w}_0 | \hat{\mathcal{M}}_{i j k l} \hat{\mathcal{M}}_{k l m n} | \mathbf{w}_0 \rangle_{m n} \quad (31)$$

where the matrix elements are given by

$$\begin{aligned}
& i j \langle \mathbf{w}_0 | \hat{\mathcal{M}}_{i j k l} \hat{\mathcal{M}}_{k l m n} | \mathbf{w}_0 \rangle_{m n} \\
&= A_{i_A j_A}^{\text{in}} \left(\frac{\partial \rho_{k_A \rightarrow l_A}}{\partial \rho_{i_A \rightarrow j_A}} \left[\left(\frac{\partial \rho_{m_A \rightarrow n_A}}{\partial \rho_{k_A \rightarrow l_A}} \right) A_{m_A n_A}^{\text{in}} + \left(\frac{\partial \varphi_{m_A \rightarrow n_B}}{\partial \rho_{k_A \rightarrow l_A}} \right) A_{m_A n_B}^{\text{out}} \right] \right. \\
&+ A_{i_A j_A}^{\text{in}} \left(\frac{\partial \varphi_{k_A \rightarrow l_B}}{\partial \rho_{i_A \rightarrow j_A}} \left[\left(\frac{\partial \rho_{m_B \rightarrow n_B}}{\partial \varphi_{k_A \rightarrow l_B}} \right) A_{m_B n_B}^{\text{in}} + \left(\frac{\partial \varphi_{m_B \rightarrow n_A}}{\partial \varphi_{k_A \rightarrow l_B}} \right) A_{m_B n_A}^{\text{out}} \right] \right. \\
&+ A_{i_B j_A}^{\text{out}} \left(\frac{\partial \rho_{k_A \rightarrow l_A}}{\partial \varphi_{i_B \rightarrow j_A}} \left[\left(\frac{\partial \rho_{m_A \rightarrow n_A}}{\partial \rho_{k_A \rightarrow l_A}} \right) A_{m_A n_A}^{\text{in}} + \left(\frac{\partial \varphi_{m_A \rightarrow n_B}}{\partial \rho_{k_A \rightarrow l_A}} \right) A_{m_A n_B}^{\text{out}} \right] \right. \\
&+ A_{i_B j_A}^{\text{out}} \left(\frac{\partial \varphi_{k_A \rightarrow l_B}}{\partial \varphi_{i_B \rightarrow j_A}} \left[\left(\frac{\partial \rho_{m_B \rightarrow n_B}}{\partial \varphi_{k_A \rightarrow l_B}} \right) A_{m_B n_B}^{\text{in}} + \left(\frac{\partial \varphi_{m_B \rightarrow n_A}}{\partial \varphi_{k_A \rightarrow l_B}} \right) A_{m_B n_A}^{\text{out}} \right] \right. \\
&+ \{A \leftrightarrow B\}, \\
&\quad (32)
\end{aligned}$$

Inserting the appropriate expressions in Eq. (25) and summing independent indices, we arrive at

$$\begin{aligned}
& \langle \mathbf{w}_0 | \hat{\mathcal{M}}^2 | \mathbf{w}_0 \rangle \\
&= \sum_{k_A} \sigma_{k_A} z_{k_A} \left[\sum_{l_A} A_{k_A l_A}^{\text{in}} \sigma_{l_A} z_{l_A} + \sum_{l_B} A_{k_A l_B}^{\text{out}} \sigma_{l_B} z_{l_B} \right] \\
&+ \sum_{k_B} \sigma_{k_B} z_{k_B} \left[\sum_{l_B} A_{k_B l_B}^{\text{in}} \sigma_{l_B} z_{l_B} + \sum_{l_A} A_{k_B l_A}^{\text{out}} \sigma_{l_A} z_{l_A} \right]. \\
&\quad (33)
\end{aligned}$$

Comparing Eq. (30) for the first order term with Eq. (33) for the second order term, we observe that instead of the in-degree $k_{k_A}^{\text{in}}$ in the first order expression, we have a sum and the corresponding adjacency matrix $A_{k_A l_A}^{\text{in}}$ (multiplied by the factors $\sigma_{l_A} z_{l_A}$) in the second order relation, which together represent exactly $k_{k_A}^{\text{in}}$ NB “steps” from k_A towards one of the neighboring nodes $l_A \in \mathcal{S}(k_A)$. The generalization of this pattern is of course precisely the NB walk (Fig. 4b) in the CI algorithm we are going to derive.

Performing the same analysis as for the previous orders, we find for the cost energy function at order $\ell = 3$,

$$\begin{aligned}
& \langle \mathbf{w}_0 | \hat{\mathcal{M}}^3 | \mathbf{w}_0 \rangle \\
&= \sum_{k_A} \sigma_{k_A} z_{k_A} \sum_{l_A} A_{k_A l_A}^{\text{in}} \left[\sum_{m_A} A_{l_A m_A}^{\text{in}} (1 - \delta_{k_A m_A}) \sigma_{m_A} z_{m_A} \right. \\
&\quad \left. + \sum_{m_B} A_{l_A m_B}^{\text{out}} \sigma_{m_B} z_{m_B} \right] \\
&+ \sum_{k_A} \sigma_{k_A} z_{k_A} \sum_{l_B} A_{k_A l_B}^{\text{out}} \left[\sum_{m_B} A_{l_B m_B}^{\text{in}} \sigma_{m_B} z_{m_B} \right. \\
&\quad \left. + \sum_{m_A} A_{l_B m_A}^{\text{out}} (1 - \delta_{k_A m_A}) \sigma_{m_A} z_{m_A} \right] \\
&+ \{A \leftrightarrow B\}, \\
&\quad (34)
\end{aligned}$$

where the factors $(1 - \delta_{k_A m_A})$ precisely capture the non-backtracking property of the walks contributing to the cost energy of a given configuration \vec{n} , in that they guarantee that the walk never returns to same node it immediately came from.

In general, when we go to higher orders $\ell \geq 4$ of the cost energy function, the NB walk may cross the same node twice and hence contain a NB loop (Fig. 4b). It is

for instance possible that a NB walk of length 3, which occurs in the cost energy function of influence at order $\ell = 4$, starts and ends in the same node. However, as shown in [18], on locally tree-like networks and for large system sizes $N = N_A + N_B$, all NB walks with loops can be neglected to leading order $\mathcal{O}(N)$.

Therefore, taking into account only the leading order contributions to the cost energy function of influence, we can finally write down the general expression for order $\ell > 1$,

$$\begin{aligned}
\langle \mathbf{w}_0 | \hat{\mathcal{M}}^\ell | \mathbf{w}_0 \rangle &= \sum_{i_A} z_{i_A} \sum_{j \in \partial \text{Ball}(i_A, \ell-1)} \left(\prod_{k \in \mathcal{P}_{\ell-1}(i_A, j)} \sigma_k \right) z_j \\
&+ \sum_{i_B} z_{i_B} \sum_{j \in \partial \text{Ball}(i_B, \ell-1)} \left(\prod_{k \in \mathcal{P}_{\ell-1}(i_B, j)} \sigma_k \right) z_j, \\
&\quad (35)
\end{aligned}$$

where $\text{Ball}(i_A, \ell)$ is the set of nodes inside a ball of radius ℓ around node i_A (Fig. 1d), with the radius defined as taking the shortest path, $\partial \text{Ball}(i_A, \ell)$ is the frontier of the ball and $\mathcal{P}_\ell(i_A, j)$ is the set of nodes belonging to the shortest path of length ℓ connecting i_A and j . Note that in the above expression the nodes j on the boundary of the ball as well as the nodes k visited during the shortest NB walk connecting i_A and j could be in either of the two modules, which is why we did not explicitly show their module label. The corresponding expression for the cost energy function to order $\ell = 1$ is given in Eq. (30).

If we agree to also consider the center node’s module label as implicit, we can write the leading order approximation of the cost energy function of influence for an arbitrary number of modules to order $\ell > 1$ as:

$$\langle \mathbf{w}_0 | \hat{\mathcal{M}}^\ell | \mathbf{w}_0 \rangle = \sum_i z_i \sum_{j \in \partial \text{Ball}(i, \ell-1)} \left(\prod_{k \in \mathcal{P}_{\ell-1}(i, j)} \sigma_k \right) z_j. \quad (36)$$

The lowest order expression for arbitrary numbers of modules is given by

$$\langle \mathbf{w}_0 | \hat{\mathcal{M}} | \mathbf{w}_0 \rangle = \sum_i \sigma_i z_i (k_i^{\text{in}} + k_i^{\text{out}}). \quad (37)$$

As stated in the beginning of this section, the problem of identifying the optimal set of influencers can be solved by minimizing the largest eigenvalue $\lambda(\vec{n})$ of the NB matrix corresponding to the NoN, which we related to the minimization of the leading order approximation of the the cost energy function of influence given by Eqs. (36) and (37). In what follows, we propose an efficient algorithm to find the minimal set of influencers.

Collective Influence algorithm for NoN, CI-NoN

Having shown that the minimal set of influencers, whose removal of input causes a breakdown of the giant mutually connected active component G , can be found

by minimizing the cost energy function of influence, we now proceed to derive the actual minimization protocol, which we call the Collective Influence algorithm.

Among all the nodes receiving an input, we want to know which node i_A or i_B in either of the two modules causes the largest drop in the cost energy function of influence when its input is removed ($n_{i_A} = 1 \rightarrow n_{i_A} = 0$) or ($n_{i_B} = 1 \rightarrow n_{i_B} = 0$).

Let us therefore briefly review the example of the simple NoN depicted in Fig.4 and answer this question for the cost energy function to order $\ell = 1$, as given in Eq.(30), assuming that all nodes initially receive an input. The important observation to be made here is that removing the input to node k_B , i.e. setting ($n_{k_B} = 1 \rightarrow n_{k_B} = 0$) affects all three state variables $\sigma_{k_B} = \sigma_{i_A} = \sigma_{j_A} = 0$ and hence decreases the cost energy function by the contribution from all of the three inactivated nodes, whereas removing the input to either node i_A or node j_A only affects their own contribution to the cost energy function. A moment's thought reveals that the crucial characteristic of node k_B , leading to such a deactivation pattern, is that both of its neighbors i_A and j_A have exactly one inter-link to k_B , i.e. their intermodule degree is precisely $k_{i_A}^{\text{out}} = 1$ and $k_{j_A}^{\text{out}} = 1$. In this case, node k_B 's input is pivotal to the activation/deactivation of its inter-link neighbors i_A and j_A .

If we formally define $\text{CI}_\ell^{\text{centric}}(i_A)$ to be the contribution to the cost energy function of influence at order $\ell + 1$ centered in i_A and proportional to σ_{i_A} , then i_A 's Collective Influence $\text{CI}_\ell(i_A)$ is the sum of its own $\text{CI}_\ell^{\text{centric}}(i_A)$ and the $\text{CI}_\ell^{\text{centric}}(j_B)$ of all nodes j_B in the other module with exactly one inter-link to i_A (Fig.1d). We call the sum of the $\text{CI}_\ell^{\text{centric}}(j_B)$ of all nodes j_B with $k_{j_B}^{\text{out}} = 1$ the *eccentric* contribution $\text{CI}_\ell^{\text{eccentric}}(i_A)$ to node i_A 's Collective Influence.

For an arbitrary number of modules, we define the Collective Influence of node i as

$$\begin{aligned} \text{CI}_{\ell=0}(i) &= z_i (k_i^{\text{in}} + k_i^{\text{out}}) + \sum_{\substack{j \in \mathcal{F}(i): \\ k_j^{\text{out}} = 1}} z_j (k_j^{\text{in}} + k_j^{\text{out}}), \\ \text{CI}_{\ell \geq 1}(i) &= z_i \sum_{j \in \partial \text{Ball}(i, \ell)} z_j + \sum_{\substack{j \in \mathcal{F}(i): \\ k_j^{\text{out}} = 1}} z_j \sum_{m \in \partial \text{Ball}(j, \ell)} z_m, \end{aligned} \quad (38)$$

where $z_i \equiv k_i^{\text{in}} + k_i^{\text{out}} - 1$. Here $\text{Ball}(i, \ell)$ is the set of nodes inside a ball of radius ℓ centered around node i (Fig.1d), with the radius defined as taking the shortest path and $\partial \text{Ball}(i, \ell)$ denotes the set of nodes residing on the frontier of the ball. We emphasize that nodes on the boundary of the ball can be in either of the modules. Indeed, the ball is grown from the central node following both intra and inter-links and thus may invade different modules of the brain NoN. Finally, we remark that the *node-eccentric* contribution to node i 's Collective Influence, given by the second term in Eq.(38), is absent in the single network case [18] and thus presents a genuine

new feature of the brain NoN.

With the Collective Influence measure (38) at our disposal, we now proceed to specify the algorithmic implementation to find and rank the minimal set of influencers ensuring global communication in the brain NoN.

The **Collective Influence algorithm** is defined as follows: Starting from the fully activated NoN, where every node is receiving an input $n_i = 1$, we progressively remove one by one the inputs ($n_i = 1 \rightarrow n_i = 0$) corresponding to the node which has the largest $\text{CI}_\ell(i)$ value (38), provided it is active $\sigma_i = 1$ (Fig.1e). After every removal of an input, the degrees of the removed node's neighbors are updated and the CI_ℓ values of the remaining active nodes are recomputed from where a new top-CI is removed and so on. The algorithm terminates when the largest active mutually connected component G is zero. The algorithm's performance increases by using larger values of the radius ℓ of the $\text{Ball}(i, \ell)$, which must however not exceed the original diameter of the NoN, for otherwise the Collective Influence is zero $\text{CI}_\ell(i) = 0$. In practice, we observe that already for $\ell = 3, 4$ the algorithm reaches the top performance (Figs.2c, d).

The Collective Influence theory developed above allows us to compute the minimal fraction q_{infl} as well as the actual configuration \bar{n}_{infl} of influencers whose removal annihilates the giant active component G and therefore brings the NoN's global communication efficiency to a halt. In the case $q < q_{\text{infl}}$, however, the giant component is nonzero, a consequence of the fact that the system of Eqs.(14) has another stable solution different from $\{\rho_{i_A \rightarrow j_A}\} = \{\rho_{i_B \rightarrow j_B}\} = \{\varphi_{i_A \rightarrow j_B}\} = \{\varphi_{i_B \rightarrow j_A}\}$ identically zero: $G = 0$. Therefore, for $q < q_{\text{infl}}$ the stability of the new solution $G(q) \neq 0$ is not controlled by the NB operator anymore, but a more complicated operator comes into play that depends on the form of the solution itself. The solution to this problem was presented in [18] and consists in implementing a reinsertion scheme. The reinsertion rule used to obtain the CI curves shown in Figs.2c, d follows the one presented in [18] and is defined as follows: given the minimal set of influencers up to q_{infl} , we reinsert one by one the inputs ($n_i = 0 \rightarrow n_i = 1$) corresponding to the node i which joins the smallest number of active clusters in the NoN when reinserted $n_i = 1$. In practice, we reinserted a finite fraction of the total number of inputs that were removed to break the giant component, before recomputing again the number of clusters the influencers to be reinserted would join. We arrive in this way to the minimal set of influencers ranked from top CI to zero. This list is then used to rank the nodes in the brain.

METHOD TO CONSTRUCT THE BRAIN NON

Dual task experiment

Our brain networks rely on functional magnetic resonance imaging (fMRI). The fMRI data consists of time-

series of the blood oxygen level dependent (BOLD) signals based on phase and amplitude response to a dual task involving visual and auditory stimuli obtained for each voxel. We use the dual-task experiment on humans explained in detail in Refs. [9, 11, 22, 27]. The data that we used in this study can be found at: <http://www-levich.engr.ccnycunyc.edu/webpage/hmakse/software-and-data>. The experiment is part of a larger neuroimaging research program headed by Denis Le Bihan and approved by the Comité Consultatif pour la Protection des Personnes dans la Recherche Biomédicale, Hôpital de Bicêtre (Le Kremlin-Bicêtre, France).

Sixteen participants (7 women and 9 men, mean age, 23, ranging from 20 to 28) performed a dual-task paradigm: a visual task of comparing an Arabic number to a fixed reference and an auditory task of judging the pitch of auditory tone. The two stimuli were applied to subjects simultaneously. Subjects were asked to press a key using right and left hand, respectively, when the number appearing on the screen was larger than a reference and the tone was high frequency.

Details of NoN reconstruction

The fMRI data we used to construct the brain NoN are taken from Ref. [22]. As outlined in great detail in Ref. [22], a 3T fMRI detector (Bruker) was utilized to record the blood oxygenation level-dependent (BOLD) signals from a T2*-weighted gradient echoplanar imaging sequence [repetition time (TR) = 1.5 s; echo time = 40 ms; angle = 90°; field of view (FOV) = 192 × 256 mm; matrix 64 × 64]. Within this setup, the entire brain was obtained in 24 slices with a thickness of 5mm each. The experimenters also recorded high-resolution images (three-dimensional gradient echo inversion-recovery sequence, inversion time = 700 mm; FOV = 192 × 256 × 256 mm; matrix = 256 × 128 × 256; slice thickness 1 mm).

Data analysis in Ref. [22], was performed with SPM2 software. In order to quantify the phase and periodicity of the fMRI data, the authors of [22], regressed the BOLD signal for each participant and trial (8 TRs of 1.5 s) against a sine and a cosine. To avoid numerical instabilities, Ref. [22] detrended the raw signal for each voxel within each trial, correcting for linear drifts and subtracting the mean (the average phase within each participant and condition was computed using the appropriate mean for circular quantities). The projections of the sine and cosine for each voxel j , are given by:

$$A^j x = \sum_i s_i \cos\left(\frac{2\pi \cdot \text{TR} \cdot i}{\text{ITI}}\right), \quad (39)$$

and

$$A^j y = \sum_i s_i \sin\left(\frac{2\pi \cdot \text{TR} \cdot i}{\text{ITI}}\right), \quad (40)$$

where $\{s_i\}$ corresponds to the detrended signal, and j denotes the voxel number. The inter-trial interval ITI was 12 sec, and TR 1.5 sec. To account for anatomical differences in brain morphology when averaging across the participants, Ref. [22] stereotactically transformed to the standardized coordinate space of Talairach and Tournoux [(Montreal Neurological Institute) MNI 152 average brain] and smoothed the regression parameters of the sine and cosine (7 mm full-width at half-maximum). As described in [27], phase and amplitude were calculated as

$$\begin{aligned} \phi^j &= \arctan(A^j y / A^j x), \\ A^j &= \sqrt{(A^j x)^2 + (A^j y)^2}, \end{aligned} \quad (41)$$

where $A^j x$ and $A^j y$ denote the regression weights of the cosine and sine for voxel j respectively. The phase was additionally multiplied by $12/2\pi$ s, in order to obtain a fraction of the stimulation period of 12 s, with a phase of 0 s indicating a peak activation coinciding with stimulus onset [22].

In order to confine the brain network reconstruction to voxels participating in the task setup, [22] estimated the fraction of the measured phases that are within the expected response range (ERR). Overall, 64 phase measurements, corresponding to four conditions per participant, were obtained. On the basis of previous characterizations of the hemodynamic response function, Ref. [22] set the ERR to the interval from 2 to 10 s, thus allowing for region-to-region and inter-condition variations. The probability for a given number of x measurements (out of the 64 total) to lie within the ERR can accordingly be calculated from the binomial distribution, as outlined in [27]. Reference [22] restricted the network analysis to voxels with more than 48 measurements within the ERR, corresponding to a binomial probability $p < 0.05$. It is worth to note that the authors of [22] evaluated the significance of the phase variations with delay using a second-level SPM model which contained all the single-trial phase measurements.

Ref. [22], performed the following two statistical tests with the collected data. First, they searched for linearly increasing phases as a function of delay (contrast $-2 - 1 \ 1 \ 2$, accounting for irregularities in the delay spacing). Second, they looked for regions with a delay by regime type interaction (contrast $1 - 1 - 11$), corresponding to a “psychological refractory period” PRP effect. Moreover, measurements of the single-trial response amplitude were tested with the same SPM model.

Definition of Brain-NoN

The construction of the 3NoN composed of AC-PPC-V1/V2 depicted in Fig. 3a consists of two main steps: first we identify the nodes belonging to each module, and then we create the intra-links and the inter-links between them (we remark that intra-links and inter-links are analogous to the strong links and weak links defined in refs [9]

and [11]). In the former step, we use the cross-correlation C_{ij} between the phases of BOLD response for each pair of voxels i and j , while in the latter step we use a machine learning algorithm to infer the pairwise interactions J_{ij} between voxels from the correlations C_{ij} . By thresholding the values of the J_{ij} we then create the connections between the voxels inside and across the modules. In the following discussion, we first explain how to identify the nodes in the three modules, and then we move to explain how we infer the connections between the nodes.

Note that the auditory cortex was activated as a major cluster in only 7 out of all 16 subjects in our percolation analysis. Fig. 5 shows the spatial location of a subject in which the auditory cortex was activated as well. While a more complete study would also include this cluster, we focused on the brain NoN composed of AC-PPC-V1/V2, which consistently appears for all 16 participants.

Detecting the modules of the brain NoN

To detect the modules in the brain NoN we first calculate the cross-correlation C_{ij} between the phases of BOLD response for each pair of voxels i and j :

$$C_{ij} = \frac{1}{N} \sum_{t=1}^N \cos(\phi_i^t - \phi_j^t). \quad (42)$$

where $N = 40$ is the number of measurements of the phases, ie, the total number that the stimulus is presented to each subject. The cross-correlation C_{ij} ranges from -1 to 1 . $C_{ij} > 0$ corresponds to positive correlations, $C_{ij} < 0$ corresponds negative correlations, and $C_{ij} = 0$ indicates the lack of correlation between a pair of voxels, i and j .

Then we use a procedure inspired by bond percolation [9, 11, 33] to separate the modules, which is described next. We progressively consider the voxels that are strongly correlated, and, by using a threshold T , we create a fictitious link between two voxels i and j if $C_{ij} > T$. At a certain percolation threshold T_c a largest connected component emerges, which gradually increases with increasing the fraction of occupied bond. Due to the modular structure of the brain, the size of the largest component increases with a series of jumps when the threshold T decreases. This growth pattern of the largest component in brain reveals that modules defined by strongly correlated connections merge one by one as T is lowered. From this observation, we can naturally identify modules in brain networks resulting from strong correlations $C_{ij} > T$ [7, 9, 11, 33]. Notice that we use this procedure only to identify which voxel belongs to which module, but we do not use the fictitious links as representative of the intra-links and inter-links. Therefore, from now on we forget about the fictitious links and we proceed by inferring the connections between voxels using a machine learning method, as explained in the next section.

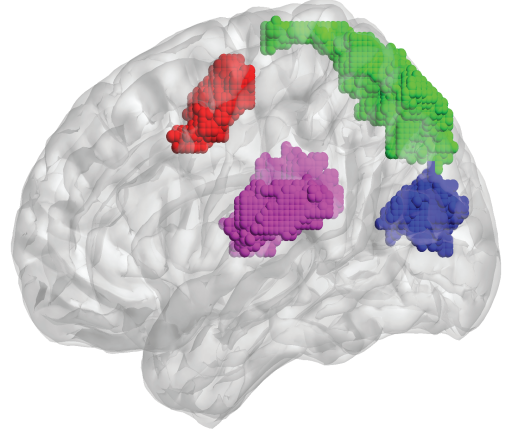


FIG. 5. Spatial location of four modules: the anterior cingulate AC (red), posterior parietal cortex PPC (green), posterior occipital visual areas V1/V2 (blue), and auditory cortex (magenta) for a typical subject. The three modules, AC, PPC, and V1/V2 appear consistently for all 16 subjects whereas the auditory cortex appears in only 7 out of all 16 participants.

Inferring the connections

To define the 3NoN composed of AC-PPC-V1/V2 depicted in Fig. 3a we reconstruct the network's intra-links and inter-links by using a Machine Learning technique called Maximum Entropy Modelling (MEM). The method has been applied to neuronal populations in [23] and it is similar to methods to infer the weights of the paths connecting two brain areas in the computational neuroscience community [24, 25]. The weight of the links that we infer are analogous to what is called direct effective connection matrix (deCM) in [25]: they embody the strength of each direct connection between points in a given brain state.

This method receives in input the set of cross-correlations $\{C_{ij}\}$ of the fMRI signals between pair of voxels measured from the fMRI BOLD response in the 3NoN, and outputs the intramodular and intermodular weights $\{J_{ij}\}$ of the path between i and j , also called interaction strengths or couplings in statistical physics. A value $J_{ij} \neq 0$ means that there exists a link between the pair of voxels i and j and the weight of this link is given by the value of J_{ij} , while if $J_{ij} = 0$ then there is no direct connection between i and j .

In order to implement the MEM, we first calculate the cross-correlation C_{ij} between the phases of BOLD response for each pair of voxels i and j as in Eq. (42). The cross-correlation C_{ij} ranges from -1 to 1 . $C_{ij} > 0$ corresponds to positive correlations, $C_{ij} < 0$ corresponds to negative correlations, and $C_{ij} = 0$ indicates the lack of correlation between a pair of voxels, i and j .

The MEM is based on the the Maximum Entropy Prin-

ciple, which implies that the most general joint distribution $P(\phi_1, \dots, \phi_N | \hat{J})$ of the phases $\phi_i \in [0, 2\pi]$, assuming solely the knowledge of the cross-correlations C_{ij} , contains only pairwise (i.e. two body) interactions (or equivalently weights) J_{ij} , and is explicitly given by the following expression:

$$P(\phi_1, \dots, \phi_N | \hat{J}) = \frac{1}{Z(\hat{J})} \prod_{i < j} e^{J_{ij} \cos(\phi_i - \phi_j)} . \quad (43)$$

The goal of this method is to estimate the interactions $\{J_{ij}\}$ such that the cross-correlations computed with the measure in Eq. (43) match the observed quantities C_{ij} , i.e.:

$$\langle \cos(\phi_i - \phi_j) \rangle \equiv \int d\vec{\phi} P(\phi_1, \dots, \phi_N | \hat{J}) \cos(\phi_i - \phi_j) = C_{ij} . \quad (44)$$

The problem of inferring the interaction matrix \hat{J} from the cross-correlation matrix \hat{C} is solved by maximizing the log-likelihood $\mathcal{L}(\hat{J} | \hat{C})$:

$$\mathcal{L}(\hat{J} | \hat{C}) = \sum_{i < j} J_{ij} C_{ij} - \log Z(\hat{J}) , \quad (45)$$

from which the inferred \hat{J}^* is obtained as:

$$\hat{J}^* = \operatorname{argmax}_{\hat{J}} \mathcal{L}(\hat{J} | \hat{C}) . \quad (46)$$

Indeed, by extremizing $\mathcal{L}(\hat{J} | \hat{C})$ with respect to J_{ij} we find

$$\begin{aligned} 0 &= \frac{\partial}{\partial J_{ij}} \mathcal{L}(\hat{J} | \hat{C}) = C_{ij} - \langle \cos(\phi_i - \phi_j) \rangle \\ &\rightarrow C_{ij} = \langle \cos(\phi_i - \phi_j) \rangle . \end{aligned} \quad (47)$$

The main difficulty of this method is to compute the quantity $\log Z(\hat{J})$, the negative of which is called free energy in statistical physics. Unfortunately there is no known closed-form for $\log Z(\hat{J})$, and, as a consequence, also to estimate the interactions J_{ij} that maximize the log-likelihood Eq. (45).

Therefore, to solve the problem, we use a Montecarlo sampling method to compute the averages $\langle \cos(\phi_i - \phi_j) \rangle$, and then we use an approximate iterative gradient ascent algorithm to update the current estimate of the couplings J_{ij} . In practice, we start from an initial guess $\{J_{ij}^0\}$ at the initial time $t = 0$ of the machine learning algorithm, and then we update the J_{ij} 's using the following rule:

$$J_{ij}^{t+1} = J_{ij}^t - \eta [\langle \cos(\phi_i - \phi_j) \rangle^t - C_{ij}] + \alpha (J_{ij}^t - J_{ij}^{t-1}) , \quad (48)$$

where the quantities $\langle \cos(\phi_i - \phi_j) \rangle^t$ are the cross-correlations computed via Montecarlo sampling using the current estimate of the couplings J_{ij}^t at time t ; η is the learning rate, and α is a damping factor that we use to help the convergence. We chose the initial $\{J_{ij}^0\}$ all equal to 0.1, the learning rate $\eta = 0.01$ and the damping factor $\alpha = 0.7$.

After estimating the couplings J_{ij} we build the 3NoN in two steps. First of all we establish the intra-links between nodes (i.e. voxels) belonging to the same module, separately for each module, and then we connect the nodes in different modules through the inter-links. Ideally we would like to put a link between two nodes i and j if and only if the corresponding J_{ij} is different from zero. However, the inference of the couplings J_{ij} is affected by noise (both because of the uncertainties in the measurements of the C_{ij} and in the Montecarlo sampling), and thus we do not have a sharp classification of zero and non-zero couplings. Therefore, we define the connections by thresholding the J_{ij} with the following criterion. First we compute the standard scores Z_{ij} of the raw couplings J_{ij} , defined as $Z_{ij} = (J_{ij} - \langle J \rangle) / \sigma$, where $\langle J \rangle$ and σ are the mean and the standard deviation of the pool $\{J_{ij}\}$. Then, for each module separately, we consider a threshold T , and we create an intra-link between two nodes in the same module if $Z_{ij} > T$.

The question of what threshold value T precisely defines the three networks is resolved using the following procedure. First we add intra-links independently in each module by choosing T to be such that the average degree $\langle k^{\text{in}} \rangle$ of intra-links is the same for each module, and equal to $\langle k^{\text{in}} \rangle = 5$.

Once the intra-links have been established, we proceed to add inter-links between pairs of voxels in different modules. Again, we consider a threshold T and we create an inter-link between two nodes i and j in two different modules if $Z_{ij} > T$. The threshold T is chosen to be such that the average $\langle k^{\text{out}} \rangle$ of the degree of the inter-links is $\langle k^{\text{out}} \rangle = 0.5$.

From this procedure we identify three predominant clusters emerging in all subjects as in previous work of dual-task data [11]: anterior cingulate (AC), posterior parietal cortex (PPC), and posterior occipital cortex (V1/V2) (Fig. 3a). The average in-degree is $\langle k^{\text{in}} \rangle = 5$ and out-degree $\langle k^{\text{out}} \rangle = 0.5$. The network data for the subject shown in Fig. 3 can be downloaded at: <http://www-levich.engr.ccny.cuny.edu/webpage/hmakse/software-and-data>.

COLLECTIVE INFLUENCE MAP OF THE BRAIN: CI-MAP

Once we construct the brain NoN, we can directly identify the location of influential nodes, through the collective influence theory. First, we compute the Collective Influence Eq. (8) in the main text for the brain NoN of each subject using $\ell = 3$. For other ℓ , we found no relevant change of the results, and increasing ℓ leads to degrading the algorithm since the networks are small and the maximum diameter is reached. We apply the adaptive CI algorithm explained in SI Text. Then, we are able to find the core nodes in the brain for a given subject according to the CI score. The typical result for the mutually connected giant component is shown in Fig. 3 for a given

subject. We identify the most influential nodes in the brain network as those obtained before the optimal percolation transition at the critical point q_{infl} . After finding the top CI voxels for each subject, we obtain the Collective Influence CI-map of the brain showing the spatial distribution of influencers, averaged over 16 subjects.

Since the number of top influencers (those included up to q_{infl}) varies with each subject (the number of nodes in the 3NoN is not the same across subjects), and to facilitate averaging across different subjects, we measure the ranking of the CI for each voxel and introduce the normalized influence by following,

$$R_{CI}(i) = \frac{r_0 - r_i - 1}{r_0}, \quad (49)$$

where r_i is the ranking of a node i and r_0 is the ranking of a baseline chosen arbitrarily. $R_{CI}(i) = 1$ corresponds to the highest CI node and $R_{CI}(i)$ decreases with decreasing r_i . In this study, we set r_0 as the ranking of top 15% node. Then, we regard the sum of $R_{CI}(i)$ as the

representative influence of a voxel i , over all subjects. In our experiments, the sum of R_{CI} ranges from 0 to 5.2 and the higher value, the more influential region.

The CI-map in Fig. 3 reveals the most influential regions in the brain during dual-task experiments. The spatial distribution of core regions predicted by CI algorithm is consistent with well-known functions of each modules as well. To be specific, the most influential regions (top CI nodes) are mainly located in the AC module which is recruited for top-down and bottom-up control. The PPC region contains a smaller portion of influential nodes next to the AC module since the PPC is responsible for both top-down and bottom-up control as well. In contrary, the influential voxels are rarely located in the V1/V2 module, which is involved in mostly processing of visual signal and bottom-up control. We conclude by saying that our theory has recently been tested in rats using pharmacogenetic interventions targeting the neural influencers responsible for memory consolidation [26].

-
- [27] Sigman M, Jobert A, Lebihan D, Dehaene S (2007) Parsing a sequence of brain activations at psychological times using fMRI. *Neuroimage* 35: 655-668.
 - [28] Wormald NC (1981) The asymptotic connectivity of labelled regular graphs. *J. Combinatorial Theory B* 31: 156-167.
 - [29] Altarelli F, Braunstein A, Dall'Asta L, Wakeling JR, Zecchina R (2014) Containing epidemic outbreaks by message-passing techniques. *Phys. Rev. X* 4: 021024.
 - [30] Krzakala F, *et al.* (2013) Spectral redemption in clustering sparse networks. *Proc. Natl Acad. Sci. USA* 110: 20935-20940.
 - [31] Mézard M, Montanari A (2009) *Information, Physics, and Computation* (Oxford University Press, USA).
 - [32] Dorogovtsev SN, Mendes JFF, Samukhin AN (2003) Metric structure of random networks. *Nucl. Phys. B* 653: 307-338.
 - [33] Eguiluz VM, Chialvo DR, Cecchi GA, Baliki M, Apkarian AV (2005) Scale-free brain functional networks. *Phys. Rev. Lett.* 94: 018102.